《生命科学前沿丛书》专家委员会

主任委员： 吴　旻

委　　员： (按汉语拼音排序)

陈永福　陈　竺　范云六　贺福初

黄大昉　李家洋　李衍达　马大龙

强伯勤　沈倍奋　王琳芳

编写人员

（按汉语拼音排序）

蔡　耘　贺福初　姜　颖

蒋代凤　刘尚义　钱小红

万晶宏　王　建　王京兰

王　妍　杨何义　应万涛

张令强　朱云平

目　录

第一章　功能基因组与蛋白质组 …… (1)
第一节　人类基因组计划及蛋白质组研究的历史背景 …… (1)
第二节　蛋白质组研究的开端及“蛋白质组”含义 …… (10)
第二章　蛋白质组研究方法 …… (24)
第一节　概述 …… (24)
第二节　大规模的蛋白质分离技术 …… (27)
第三节　高通量蛋白质鉴定技术 …… (39)
第三章　二维电泳的蛋白质提取与样品制备 …… (48)
第一节　细胞裂解的方法 …… (49)
第二节　蛋白质的分步提取技术 …… (57)
第三节　亚细胞分离与蛋白质提取 …… (59)
第四章　二维电泳与细胞蛋白质的分离 …… (68)
第一节　概述 …… (68)
第二节　一维等电聚焦电泳 …… (72)
第三节　二维 SDS-聚丙烯酰胺凝胶电泳 …… (80)
第四节　胶上蛋白的检测 …… (81)
第五节　存在问题 …… (87)
第五章　图像分析与细胞蛋白谱的建立 …… (91)
第一节　蛋白电泳图像分析系统 …… (91)
第二节　二维电泳蛋白谱数据库 …… (97)
第六章　生物质谱技术与蛋白质鉴定 …… (101)
第一节　基质辅助激光解吸电离飞行时间质谱 …… (101)
第二节　液相色谱-电喷雾-串联质谱 …… (105)
第三节　肽质量指纹谱鉴定蛋白质技术 …… (110)
第四节　串联质谱数据鉴定蛋白质技术 …… (116)
第七章　蛋白质翻译后修饰的鉴定 …… (130)
第一节　磷酸化蛋白质的鉴定 …… (130)
第二节　糖基化蛋白质的鉴定 …… (145)

第八章　定量蛋白质组学研究技术 …… (165)
第一节　利用荧光染料进行定量的蛋白质组分析技术 …… (165)
第二节　运用质谱进行定量的蛋白质组分析技术 …… (168)
第三节　蛋白质芯片技术 …… (171)
第九章　蛋白质组研究中的生物信息学 …… (174)
第一节　生物信息学简介 …… (174)
第二节　数据库的构建 …… (180)
第三节　蛋白质组研究中常用的网站及数据库 …… (188)
第十章　细胞器与蛋白质复合体的组成分析 …… (203)
第一节　细胞器的分离 …… (203)
第二节　细胞器的组成分析 …… (214)
第三节　蛋白复合体分离及组成分析 …… (217)
第十一章　蛋白质间连锁图的建立 …… (242)
第一节　酵母双杂交技术及其应用 …… (242)
第二节　免疫共沉淀技术 …… (251)
第三节　细胞共定位技术 …… (257)
第十二章　多能干细胞定向分化的蛋白质组研究 …… (266)
第十三章　肿瘤发生与发展的比较蛋白组研究 …… (285)
第一节　蛋白质组与肿瘤 …… (285)
第二节　肺癌转移的蛋白质组研究 …… (285)
第三节　辐射致癌相关蛋白质分子标志物的研究 …… (296)
第十四章　细胞凋亡的蛋白质组研究 …… (304)
第一节　细胞凋亡的信号传导 …… (304)
第二节　细胞凋亡的蛋白质组研究 …… (315)
第十五章　蛋白质组与新药开发 …… (330)
第一节　药物分子靶标的高通量筛选体系 …… (332)
第二节　药物的高通量筛选体系 …… (337)
第三节　药物作用的监测评价体系 …… (340)

第一章　功能基因组与蛋白质组

基因研究是20世纪生命科学的主线。20世纪的上半叶，以遗传学为代表，生命科学通过对基因分离、独立分配、连锁及化学属性等的研究，最后以作为遗传信息载体的DNA双螺旋结构的提出而告捷；20世纪的下半叶，以分子生物学为代表，生命科学通过对基因复制、转录、翻译及遗传密码的分析与破译，最终以统一生命世界各层次、生命科学各分支的“中心法则”的问世而集成；20世纪90年代，随着全球性基因组计划尤其是人类基因组计划（HGP）规模空前、速度惊人的推进，基因研究已接近“登峰造极”，人类对生命世界的理性认识达到了前所未有的深度与广度。

人类基因组计划被誉为20世纪的三大科技工程之一。其划时代的研究成果——人类基因组序列草图的完成，宣告了一个新的纪元——“后基因组时代”的到来。其中，功能基因组学（functional genomics）成为研究的重心，蛋白质组学（proteomics）则是其“中流砥柱”。正因为如此，*Nature*、*Science*分别在2001年2月15日、16日公布人类基因组草图的同时，分别发表了“And now for the proteome”（Nature 409：747，2001）、“Proteomics in genomeland”（Science 291：1221，2001）的述评与展望，将蛋白质组学的地位提到前所未有的高度，认为是功能基因组学这一前沿研究的战略制高点，蛋白质组学将成为新世纪最大战略资源——人类基因尤其是重要功能基因争夺战的重要“战场”。

人们在欢呼基因组计划辉煌业绩之时，亦愈来愈清醒地意识到一项更艰巨、更宏大的任务即基因组功能的阐明已经摆在面前，生命科学几乎在转瞬之间开始了新的征程——蛋白质组研究[1-3]，进入了一个新的纪元——后基因组时代（postgenome era）[4,5]。人类经过一个世纪的跋涉，重返现代生命科学的发源地之一，蛋白质——这一生命活动的执行体。当然，这不是简单的回归，而是一次真正的黑格尔式的“重返”。

第一节　人类基因组计划及蛋白质组研究的历史背景

一、基因组计划的成就

以人类基因组计划为代表的基因组计划是20世纪仅次于曼哈顿原子弹研制计划与阿波罗登月计划的重大科技工程。其中，HGP旨在完成人基因组24条染

色体上5万左右基因的作图（遗传图与物理图）和30亿碱基的DNA全序列的测定。此计划自1990年实施以来进展神速：1994年人基因组全套遗传连锁图发表[6]，1995年全基因组覆盖率高达94%的物理图问世[7]；同年，汇集了人基因组初步全物理图，3、12、16、22号染色体高密度物理图以及30余万左右cDNA（EST）序列信息的“人基因组指南”经 *Nature* 结集出版[8]；2000年6月26日，宣告人类基因组序列草图测定完成；2001年2月15日、16日，国际人类基因组计划与美国Celera公司分别在 *Nature*、*Science* 公布人类基因组草图[9]。与此同时，模式生物与致病微生物等的基因组研究亦如火如荼地展开。自1995年支原体（*Mycoplasma genitalium*）和流感嗜血杆菌（*Hemophilus influenzae*）基因组全序列发表[10,11]以来，已相继有10余种原核生物的基因组全序列发表，如大肠杆菌（K-12）[12]。更令人振奋的是，第一个真核生物——酵母的基因组全序列于1996年完成[13]，次年，Nature再次推出（酵母基因组指南）专辑[14]。此外，多细胞真核生物线虫（*Caenorhabditis elegans*）的全基因组序列测定也取得长足进展[15]。截至2001年底，已有不少于75种生物的基因组全序列测定完成，基因组计划已经进入全面收获的“金秋时节”（表1-1综合了共75个目前已经完成全基因组序列测定的生物），而“海量”的基因序列数据为生命科学多层次、多分支的研究提供了丰富的宝藏。

表1-1　目前已经完成全基因组序列测定的生物

（共75个，截至2002年初）

生物种类	基因组大小/kb	ORF数目	测序机构
古细菌（12个）			
Methanococcus jannaschii DSM 2661	1664	1750	美国伊利诺伊州立大学和TIGR公司
Methanobacterium thermoautotrophicum delta H	1751	1918	基因组治疗公司和俄亥俄州立大学
Archaeoglobus fulgidus DSM4304	2178	2493	美国伊利诺伊州立大学和TIGR公司
Pyrococcus horikoshii（*shinkaj*）OT3	1738	1979	东京大学和日本NITE（National Institute of Technology and Evaluation）
Aeropyrum pernix K1	1669	2620	日本NITE（National Institute of Technology and Evaluation）
Pyrococcus abyssi GE5	1765	1765	法国Genoscope中心

续表

生物种类	基因组大小/kb	ORF数目	测序机构
Halobacterium sp. NRC-1	2014	2058	华盛顿大学和麻省州立大学
Thermoplasma acidophilum	1564	1478	德国 Max Planck 生化研究所和 Medigenomix 公司
Thermoplasma volcanium GSS1	1584	1524	日本 AIST（National Institute of Advanced Industrial Science and Technology）
Sulfolobus solfataricus P2	2992	2977	欧盟和加拿大生物信息资源中心
Sulfolobus tokodaii 7	2694	2826	日本 NITE（National Institute of Technology and Evaluation）
Pyrobaculum aerophilum IM2	2222	2587	加利福尼亚理工学院（California Institute of Technology）和加州大学
细菌（57个）			
Haemophilus influenzae KW20	1830	1850	TIGR 公司
Mycoplasma genitalium G-37	580	468	TIGR 公司
Synechocystis sp. PCC6803	3573	3168	日本 Kazusa DNA Research Institute (KDRI)
Mycoplasma pneumoniae M129	816	677	德国海德堡大学
Escherichia coli K12- MG1655	4639	4289	美国威斯康辛大学
Helicobacter pylori 26695	1667	1590	TIGR 公司
Bacillus subtilis 168	4214	4099	法国 Pasteur 研究所和日本京都大学
Borrelia burgdorferi B31	1230	1256	Brookhaven Natl 实验室和 TIGR 公司
Aquifex aeolicus VF5	1551	1544	美国伊利诺伊州立大学
Mycobacterium tuberculosis H37Rv (lab strain)	4411	4402	英国 Sanger 中心
Treponema pallidum subsp. pallidum Nichols	1138	1041	美国德克萨斯州和 TIGR 公司

续表

生物种类	基因组大小/kb	ORF 数目	测序机构
Chlamydia trachomatis serovar D	1042	896	美国斯坦福大学和加利福尼亚大学伯克莱分校
Rickettsia prowazekii Madrid E	1111	834	瑞典 Uppsala 大学
Helicobacter pylori J99	1643	1495	基因组治疗公司和美国 Astra 公司
Chlamydia pneumoniae CWL029	1230	1052	美国斯坦福大学和加利福尼亚大学伯克莱分校
Thermotoga maritima MSB8	1860	1877	TIGR 公司
Deinococcus radiodurans R1	3284	3187	TIGR 公司
Ureaplasma urealyticum serovar 3	751	650	Eli Lilly 公司和美国 Alabama 大学、Perkin Elmer 公司
Campylobacter jejuni NCTC 11168	1641	1654	英国 Sanger 中心和英国 Leicester 大学、伦敦卫生与热带医学学校
Chlamydia pneumoniae AR39	1229	1052	英国 Sanger 中心和加拿大 Manitoba 大学
Chlamydia trachomatis MoPn Nigg	1069	924	英国 Sanger 中心和加拿大 Manitoba 大学
Neisseria meningitidis MC58 (serogroup B)	2272	2158	TIGR 公司
Neisseria meningitidis Z2491 (serogroup A)	2184	2121	英国 Sanger 中心和牛津大学、德国 Max-Planck 分子遗传学研究所
Bacillus halodurans C-125	4202	4066	日本海洋科学技术中心

续表

生物种类	基因组大小/kb	ORF 数目	测序机构
Chlamydia pneumoniae J138	1228	1070	日本 Yamaguchi 大学和 KYUSHU 大学
Xylella fastidiosa CVC 8.1.b clone 9.a. 5.c	2679	2904	巴西 ONSA 中心
Vibrio cholerae serotype O1, Biotype El Tor, strain N16961	4000	3885	TIGR 公司
Pseudomonas aeruginosa PAO1	6264	5570	美国华盛顿大学和 Chiron 公司
Buchnera sp. APS	640	564	日本东京大学和 RIKEN 研究所（The Institute of Physical and Chemical Research）
Mesorhizobium loti MAFF303099	7596	6752	日本 Kazusa DNA Research Institute（KDRI）
Escherichia coli O157: H7 EDL933	4100	5283	美国威斯康辛大学
Mycobacterium leprae TN	3268	1604	英国 Sanger 中心和法国 Pasteur 研究所
Escherichia coli O157: H7. Sakai	5594	5448	日本京都大学
Pasteurella multocida Pm70	2250	2014	美国明尼苏达州立大学
Caulobacter crescentus	4016	3737	TIGR 公司

续表

生物种类	基因组大小/kb	ORF 数目	测序机构
Streptococcus pyogenes SF370（M1）	1852	1696	美国俄克拉荷马州立大学
Lactococcus lactis IL1403	2365	2266	法国 Genoscope 中心
Staphylococcus aureus N315	2813	2594	日本 NITE（National Institute of Technology and Evaluation）和 Juntendo 大学、Tsukuba 大学、东京大学、KYUSHU 大学等
Staphylococcus aureus Mu50	2878	2697	日本 NITE（National Institute of Technology and Evaluation）和 Juntendo 大学、Tsukuba 大学、东京大学、KYUSHU 大学等
Mycobacterium tuberculosis CDC 1551	4403	4187	TIGR 公司
Mycoplasma pulmonis	963	782	法国 Genoscope 中心
Streptococcus pneumoniae TIGR4	2160	2094	基因组治疗公司
Sinorhizobium meliloti 1021	6690	6205	欧盟和美国斯坦福大学
Streptococcus pneumoniae R6	2038	2043	Eli Lilly 公司
Rickettsia conorii Malish 7	1268	1374	法国 Genoscope 中心
Yersinia pestis CO-92 Biovar Orientalis	4653	4012	英国 Sanger 中心和 MDS 公司、Dstl 实验室、帝国学院
Salmonella typhi CT18	4809	4600	英国 Sanger 中心和帝国学院

续表

生物种类	基因组大小/kb	ORF 数目	测序机构
Salmonella typhimurium, LT2 SGSC1412	4857	4597	美国华盛顿大学
Listeria innocua Clip11262, rhamnose-negative	3011	2981	法国 Pasteur 研究所
Listeria monocytogenes EGD-e	2944	2855	法国 Pasteur 研究所
Nostoc sp. PCC 7120	6413	5366	日本 Kazusa DNA Research Institute (KDRI) 和美国密西根州立大学
Agrobacterium tumefaciens C58-Cereon	4915	5299	美国 Monsanto 公司和 Cereon 公司、Richmond 大学
Agrobacterium tumefaciens C58-DuPont	4915	5402	美国华盛顿大学和 DuPont 公司、巴西 Campinas 大学
Ralstonia solanacearum GM11000	5810	5120	法国 Genoscope 中心和 INRA 公司、CNRS 公司
Brucella melitensis 16M	3294	3197	美国 Scranton 大学和 Integrated Genomics 公司
Clostridium perfringens 13	3031	2660	日本 Tsukuba 大学和 Kitasato 大学、Kyushu 大学
真核生物（9 个）			
Saccharomyces cerevisiae S288C	12 069	6294	国际合作
Caenorhabditis elegans	97 000	19 099	美国华盛顿大学和英国 Sanger 中心
Drosophila melanogaster	137 000	14 100	美国 Celera 公司和加大伯克莱分校 DGP (Drosophila Genome Project)、Baylor College of Medicine、欧洲 DGP

续表

生物种类	基因组大小/kb	ORF 数目	测序机构
Arabidopsis thaliana	115 428	25 498	国际合作
Homo sapiens	约 3 000 000	35 000～50 000	国际合作
Guillardia theta	551	464	加拿大大不列颠哥伦比亚大学和 Canadian Institute for Advanced Research (CIAR)、德国 Philipps 大学
Leishmania major Friedlin Chromosome 1	257	79	美国 Seattle Biomedical Research Institute (SBRI)
Plasmodium falciparum 3D7 Chromosome 2	947	205	TIGR 公司
Plasmodium falciparum 3D7 Chromosome 3	1060	220	英国 Sanger 中心

二、基因组计划的局限

任一生物基因组计划［此处按经典含义指结构基因组学（structural genomics）分析］的完成均标志着三套完整数据的获得：遗传图、物理图、全序列图。理论上，这三套数据将提供此生物所有基因在染色体上的精确定位、基因内部序列结构与所有基因间隔序列。但是，由于真核生物中基因结构的复杂性以及现有基因识别（gene identification）理论与技术发展的严重不足，此情况只适用于原核生物或低等真核生物。正因为如此，即使 HGP 能在 2001 年完成，也并不表明此时人类对自身基因组的所有基因及其间隔序列已完全确定。真核生物尤其是高等真核生物已测定基因组中 ORF（可读框，或名开放阅读框架）的确定仍是未解决的重大问题，而一个基因在 ORF 确定前很难从分子水平上进行实质性的功能分析。

基因调控研究表明，即使是简单的微生物（如大肠杆菌），其基因组的所有基因也不同时表达。通常情况下，生物的基因组只表达少部分基因，而且表达的基因类型及其表达程度随生物生存环境及内在状态的变化而表现极大的差别，且

此差别存在严格调控的时空特异性。基因组计划即使已确定某生物基因组内的全部基因，也不能告诉人们哪些基因在何时何地以何种程度表达，而生命过程的精确机制很大程度上正是基于这类基因的精细调控。为了弥补基因组计划这一天然的局限，近年人们相继引进一系列大规模基因表达检测技术，如微阵列法（microarray）[16]、DNA 芯片（DNA chips）[17]及 SAGE（serial analysis of gene expression）[18]等。这些方法虽然能够定性、定量且大规模地检测基因的表达产物 mRNA，但 mRNA 由于自身存在贮存、转运、降解、翻译调控及产物的翻译后加工，难以准确地反映基因的最终产物/基因功能的真正执行体——蛋白质的质与量。

基因与其编码产物蛋白的线性对应关系只存在于新生肽链而不是最终的功能蛋白中。30 多年前，人们即已普遍发现新生肽链合成后存在多种加工、修饰过程；更有甚者，近些年来人们发现蛋白质间亦存在类似于 mRNA 分子内的剪切、拼接，并证明其基本元件“intein”广泛存在于多种蛋白质中[19]。此类过程的存在无疑进一步扩大了基因编码的蛋白质与其最终的功能蛋白间所存在的序列差距。而大量蛋白尤其是重要调控蛋白的化学修饰（如糖基化、磷酸化）、剪切加工（如酶原降解、结构域拼接）不但可改变其立体结构，而且是实施其功能与调节的重要结构基础。这些均不能从其基因编码序列中预测，而只能通过对其最终的功能蛋白进行分析。

从上可见，基因虽是遗传信息的源头，而功能性蛋白是基因功能的执行体。基因组计划的实现固然为生物有机体全体基因序列的确定，为未来生命科学研究奠定了坚实的基础，但是它并不能提供认识各种生命活动直接的分子基础，其间必须研究生命活动的执行体——蛋白质这一重要环节。

任一层次的生命活动均是非线性复杂系统中各种功能单元协同、整合的结果，生命活动的最小单元——“细胞”即是多类“蛋白机器”（protein machine）的有机组合[20]。人类对于蛋白质的研究已逾百年，但以往的视角只是针对生命活动中某一种或某几种蛋白质，这样难以形成一种整体观，难以系统透彻地阐释生命活动的基本机制。因此，无论是从基因组计划的局限、还是从蛋白质研究的自身发展而言，大规模、全方位的蛋白质研究均是势在必行。

蛋白质是生物细胞赖以生存的各种代谢和调控途径的主要执行者，因此蛋白质不仅是多种致病因子对机体作用最重要的靶分子，并且也成为大多数药物的靶标乃至直接的药物。药靶，来源于对生命活动的生理病理过程的研究；药靶，又形成制药业的发展源头。蛋白质组学正是近年来新发展起来的强有力的发现药靶的技术平台，作为一个新的学科发展领域，它对所有及时进入的国家都将提供巨大的机会。机不可失，时不我待。

一项科学统计表明：在 20 世纪 90 年代中期，全世界制药业用于找寻新药的药靶共约 483 个，它们主要是蛋白质（受体占 45%，酶占 28% 等）；而当时全世

界正在使用的药物总数约是2000种，其中85%都是针对上述483种药靶。这483种药靶分子构成了全世界药厂的最重要的发展源泉。从功能基因组学的角度，人们认为每种疾病平均与10个左右基因相关，而每种基因又与3～10种蛋白质相关，如果以人类主要的100～150种疾病进行计算，则应该有3000～15 000种蛋白质具有成为药靶的可能。也就是说还可能有几千到上万种的新药靶将被发现，这将是功能基因组研究有可能带来的一笔巨大的科学、经济财富；不容置疑，这也是为什么蛋白质组学作为发现药靶的主要技术平台在20世纪90年代末期以来越来越受国际巨型跨国制药集团垂青的重要原因所在。

三、蛋白质研究技术方法的突破

蛋白质的研究在20世纪70年代以前一直优于核酸。其后，由于DNA重组、测序、PCR等新方法的不断涌现，核酸研究后来居上，并远远超出而成为生命科学的主导，但蛋白质研究尤其是相关技术的发展并未停滞不前。其中O'Farrel PH于1975年建立的二维电泳（又名双向电泳，2-DE）技术使蛋白分辨达到成千上万种，因而完全可以用于组织与细胞中大规模蛋白质的分离[21]；近年开发的多种图像分析系统与软件以及大规模样品处理系统更使其如虎添翼。20世纪80年代末期Hillenkamp F发展的激光解吸质谱、Fenn J设计的电喷雾质谱可以高效、精确地测量生物大分子的质量并测定部分序列[22]，进而用于数据库的检索；Mann M等则在此基础上通过建立“肽质量指纹图谱与肽序列标签”等技术，实现了质谱准确、快速、自动化、大规模鉴定蛋白质的飞跃[23]。

第二节　蛋白质组研究的开端及“蛋白质组”含义

一、蛋白质组研究的开端

“proteome”（蛋白质组）一词由Marc Wilkins于1994年在意大利Siena的一次2-DE电泳会议上首次提出。其导师澳大利亚Macquarie大学的Keith Williams于同年向澳政府提出一项建议：通过对某一种生物的所有蛋白质全部进行质谱筛选与序列分析，以一种不同于DNA快速测序的途径对其提供分子水平的全面分析。1995年，悉尼大学Humphery Smith I实验室与Williams等4家实验室合作，对至今已知最小的自我复制生物 *Mycoplasma genitalium*（一种支原体）进行了蛋白质成分的大规模分离与鉴定，并在文献中首次公开使用“proteome”一词，同时指出该文所采用的技术体系对于大规模鉴定并分析基因对应的产物以及发现新型蛋白均具有十分重要的意义[1]。

二、蛋白质组的含义

根据 Wilkins MR 等[24]的定义，“proteome”一词源于“PROTEin”与“genOME”的杂合，意指“一种基因组所表达的全套蛋白质”；Swinbanks[3]则指出“proteome”代表一完整生物的全套蛋白质。与此同时，Kahn P 则认为“proteome”反映不同细胞的不同蛋白质组合[2]。由此可见，“proteome”有三种不同的含义：一个基因组、一种生物或一种细胞/组织所表达的全套蛋白质。

三、蛋白质组研究的技术路线与相关技术

一般细胞含有数千种乃至上万种蛋白质。蛋白质组研究的宗旨是将组织或细胞所有蛋白质（至少是大部分）分离与鉴定。为达到目的，它引进了下列技术[21]：双向电泳（2-DE），如 ISO-DALT、IPG-DALT 或 NEPHGE；图像分析系统，如 ELSIE 4 & 8、gellab Ⅰ & Ⅱ、MELANIE Ⅰ & Ⅱ、QUEST Ⅰ & Ⅱ与 PDQUEST、TYCHO & KEPLAR；蛋白质鉴定方法，如氨基酸组成分析、序列测定、肽质量指纹图（peptide mass fingerprinting，PMF）、相对分子质量精确测定；HTS（high throughput system）系统与大规模样品处理机器人；数据库设置与检索系统[25]。其中，蛋白质鉴定采用了新近出现的新型质谱（MS）技术，如 MALDI-TOF（matrix assisted laser desorption ionisation-time of flight）MS 与 ESI（electrospray ionisation）/MS/MS。

大规模蛋白质组分析过程包括样品制备、图像分析、蛋白质成分的分析与鉴定。其技术路线及数据处理见有关章节。其中“层次分析”（hierarchial analysis）包括[26]：氨基酸分析、肽质量指纹图、氨基酸分析与 PMF 联合、序列标签途径、N 端 Edman 降解蛋白与微量测序、蛋白质内肽微量测序、MS（MALDI-TOF，ESI）微量测序、“Ladder”测序、MS 对 PVDF 膜或电泳胶上低拷贝分子的系统筛选、基因组文库中克隆片段的倾向性表达。

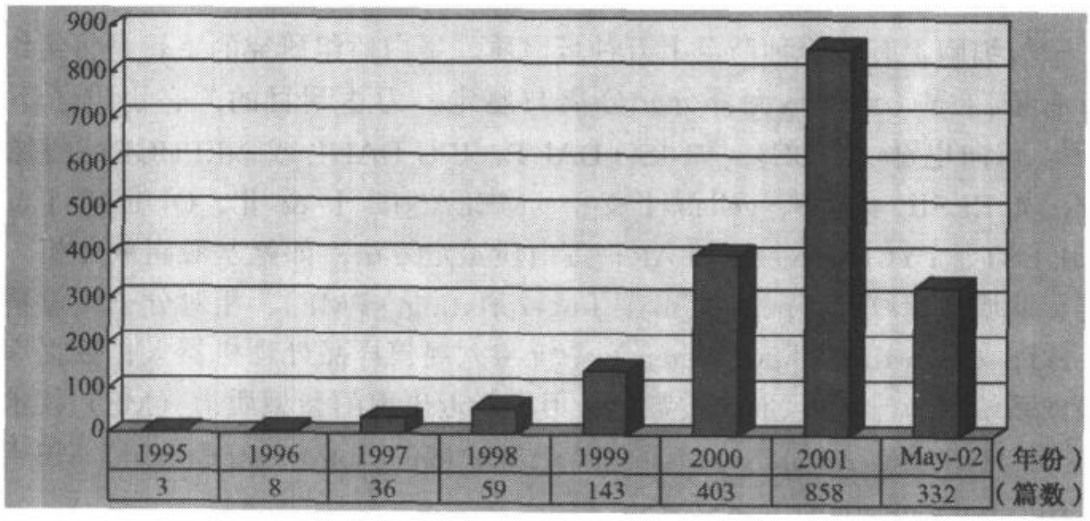
四、蛋白质组研究的国内外现状

1. 发展速度与规模

自 1995 年“proteome”一词问世以来，截至 1997 年底相关文献已达 41 篇，其中研究论文 28 篇，各类评述综述 13 篇。研究论文中，1995 年 1 篇，1996 年 4 篇，1997 年 23 篇，近 4 年更是以指数式增长。其论文增长速度见表 1-2。1995 年首倡“proteome”的两家澳大利亚实验室，1997 年分别挂牌“Centre for Proteome Research and Gene Product Mapping, National Innovation Centre”和“Aus-

tralia Proteome Analysis Facility"。同年，丹麦成立"Centre for Proteome Analysis in Life Sciences"，美国成立专事此类研究与开发的两家公司："Proteome, Inc. Beverly, MA"和"Large Scale Biology Corp, Rockville, MD"。参与的国家，1995年只有澳大利亚，1997年则发展到美国、丹麦、瑞士、英国、法国、日本、瑞典、意大利、德国等10国。国际著名学府如哈佛、斯坦福、耶鲁、密执安、华盛顿大学、欧洲分子生物学实验室、巴斯德研究所、瑞士联邦工业学院等均挤身此类研究。其中，澳大利亚悉尼大学与Macquarie大学发展势头强劲，但美欧多家实验室已奋起直追，正是"群雄纷起，逐鹿中原"。

表1-2　蛋白质组研究论文增长速度



2. 研究材料

1995年，Wasinger等[1]在第一篇蛋白质组研究文章中研究的对象为目前已知最小但能自主复制的原核微生物——支原体 *Mycoplasma genitalium*。1996年，研究对象即扩展到单细胞真核生物——酵母[27]以及人体正常组织及病理标本[28]，进而突破了早期人们普遍认为的"蛋白质组研究只适用于基因组计划已完成的生物"的界限[1,24]。因此，1997年，研究对象一下扩展到14种生物，其中虽然绝大多数为原核生物，但也包含多细胞真核生物如线虫[29]。近4年来，蛋白质组研究对象已无任何限制：无需基因组计划完成（当然完成者更好），无原核生物/真核生物、单细胞/多细胞、组织之分。

3. 研究范围

"蛋白质组"不仅其研究已成为具有重大战略意义的科学命题，而且其分析已成为一种十分有效且应用广泛的研究手段。正因如此，蛋白质组的研究与分析

其范围在短时间内即扩展到令人惊诧的程度。据不完全统计，目前至少已涉及如下方面：① 蛋白质，如蛋白质组作图[30]、蛋白质组成分鉴定[31]、蛋白质组数据库构建[32]、新型蛋白质发掘[33]、蛋白质差异显示[34]、同工体（isoform）比较[35,36]；② 基因，如功能基因组计划[35~37]、基因产物识别[38]、基因功能鉴定[39]、基因调控机制分析[40,41]；③ 重要生命活动的分子机制，包括细胞周期[26]、细胞分化与发育[29,32]、肿瘤发生与发展[42,43]、环境反应与调节[32,37,41]、物种进化等[44~46]；④ 医药靶分子寻找与分析，靶分子类型包括新型药物靶分子[33,34,44]、肿瘤恶性标志[42,43]、人体病理介导分子[28]、病原菌毒性成分[33,44]。由此不难看到，“蛋白质组”研究与分析已涉足生命科学中一系列热点领域。

蛋白质组作图是早期蛋白组研究的主要领域。经过最初二年的努力，在1995年已测定10种蛋白质组图（2-DE）[24]的基础上，又测定了11种生物或组织的蛋白质组图，其中3种生物（线虫[29]、豆科植物根瘤菌属[34]、*Ochrobactrum anthropi*[47]）的蛋白质组图超过1600点，3种蛋白质组图（大肠杆菌[39]、盘基网柄菌[32]、人正常组织与病理组织[28]）建立数据库并上互联网；各类生物中第一个完整的蛋白质组数据库（YPD）完成[48]，含6021种蛋白；提出“蛋白差异显示”概念，并用于环境应激、基因突变、病理进程等研究[34]；建立“proteomic contig”方法，进而使蛋白质组图分辨率提高10倍[47]；提出“蛋白连锁图”（protein linkage map）概念，改进双杂交系统，用于蛋白质组相互作用网络的分析[49,50]；联合液体自动取样器与LC/ES-MS技术，使蛋白质鉴定速度达20点/日[35]；建立2-DE中糖蛋白与膜蛋白微量鉴定方法[51,52]；建立“Streamlined样品处理”胶转膜技术，突破了大规模蛋白测序与氨基酸组成分析这两个严重影响蛋白质组分析中蛋白质鉴定的限速步[53]。

4. 国内外研究现状

国外研究现状

基因、尤其是基因组研究形成了20世纪生命科学研究一道亮丽的风景线。根据2001年5月23日所公布的结果，已公开发表67个生物体的基因组全序列（含染色体），其中包括古细菌11个，细菌41个，真核生物15个，此外有208个原核生物和152个真核生物或其染色体的基因组正在测序（http://igweb.integratedgenomics.com/GOLD/）。基因组研究所产生的“海量”数据和大量新型技术以前所未有的深度与广度极大地推动了生物医学多学科的飞速发展。

但是，随着大量生物体全基因组序列的获得、特别是人类基因组序列草图的完成，基因组研究的战略重点不可避免地从结构基因组学（structural genomics）转向功能基因组学（functional genomics），而蛋白质组学（proteomics）正是作为功能基因组研究的重要支柱在20世纪90年代中期应运而生[1]。也正因如此，*Nature*、*Science*在2001年2月公布人类基因组草图的同时，分别发表了“And

now for the proteome"[54]、"Proteomics in genomeland"[55]的述评与展望，将蛋白质组学的地位提到前所未有的高度，认为是功能基因组学这一前沿研究的战略制高点，蛋白质组学将成为新世纪最大战略资源——人类基因争夺战的重要"战场"。近年，*Nature*、*Science* 和 *Cell* 以及其他一些重要杂志接连刊登有关蛋白质组学的评论文章或原始论著，明确表明了蛋白质组学已经成为新世纪生命科学研究的前沿[56-79]。

蛋白质组学的第一篇原始论著 1995 年发表于国际上并不著名的 *Electrophoresis*[1]。如果说蛋白质组学刚诞生时没有立即得到国际生命科学界主流的高度重视，那么近三年已发生巨大的变化。美国国立卫生研究院（NIH）所属的国立肿瘤研究所（NCI）投入了大量经费支持蛋白质组研究，其中一千万美元用于在密执安医学院建立一个有关肺、直肠、乳腺和卵巢等肿瘤的蛋白质组数据库；此外，国立肿瘤研究所和美国食品与药物管理局联合，投入数百万美元，资助建立一个有关癌症不同发病阶段和治疗阶段的蛋白质组数据库。美国能源部不久前也启动了蛋白质组项目，旨在研究涉及环境和能源的微生物和低等生物的蛋白质组。欧共体目前正资助酵母蛋白质组研究。英国生物技术和生物科学研究委员会最近也资助了三个研究中心，对一些已完成或即将完成全基因组测序的生物进行蛋白质组研究。在法国，新成立了五个区域性遗传基地（genopoles），它们将得到为期三年的资助，每年约为五百万美元，这些经费将平均用于基因组、转录组（transcriptome）和蛋白质组研究。德国的联邦研究部提供七百多万美元，在东德的 Rostock 建立了一个蛋白质组学中心。1997 年澳大利亚政府即着手建立第一个全国性的蛋白质组研究网 APAF（Australian Proteome Analysis Facility）。APAF 为该国的有关实验室提供一流的仪器设备，并把它们整合在一起进行大规模的蛋白质组研究。日本的科学与技术委员会也已先期由政府出资三百万美元开展蛋白质组研究。由此可见，蛋白质组学虽然问世不到 10 年，但鉴于其战略的重要性和技术的先进性，西方主要发达国家在这一新型领域争先恐后、均已投入巨资全面启动此领域的研究。

更有甚者，由于蛋白质组学研究比基因组学研究更接近实用，具有着巨大的市场前景，企业与制药公司纷纷斥巨资开展蛋白质组研究。如独立完成人类基因组测序的 Celera 公司已宣布投资上亿美元于此领域[78]；又如日内瓦蛋白质组公司与布鲁克质谱仪制造公司联合成立了国际上最大的蛋白质组研究中心。不难看到，蛋白质组学已成为西方各主要发达国家、各跨国制药集团竞相投入的"热点"与"焦点"。

蛋白质组学的前沿大致分为三大方向：① 针对有基因组或转录组数据库的生物体或组织/细胞，建立其蛋白质组或亚蛋白质组（或蛋白质表达谱）及其蛋白质组连锁群，即 compositional proteomics 研究；② 以重要生命过程或人类重大疾病为对象，进行重要生理/病理体系或过程的比较蛋白质组学研究，即 compar-

ative proteomics 研究；③ 蛋白质组学支撑技术平台和生物信息学的研究。

国外大部分蛋白质组表达谱的研究论文发表于 2000 年下半年以后，且大多建立在已完成基因表达谱的基础上，表明目前在基因组或转录组基础上开展蛋白质组表达谱的研究是一个新的方向[79]。人类重大疾病的蛋白质组研究通常采用比较蛋白质组分析方法。近年来，蛋白质组学技术在研究细胞的增殖、分化、异常转化、肿瘤形成等方面进行了有力的探索，涉及到白血病、乳腺癌、结肠癌、膀胱癌、前列腺癌、肺癌、肾癌和神经母细胞瘤等，鉴定了一批肿瘤相关蛋白，为肿瘤的早期诊断、药靶的发现、疗效判断和预后提供了重要依据[80]。高通量、高灵敏度和规模化的双向凝胶电泳-质谱是目前最流行和较可靠的技术平台[81,82]；酵母双杂交技术已被用于研究蛋白质连锁群和蛋白质功能网络系统，且分别发表于 *Nature* 与 *Science*[83, 84]。生物信息学方法在蛋白质组学研究领域亦得到有效的利用，其中突出的代表是 Eisenberg D 等联合采用 phylogenetic profiles 法、Rosetta stone 法和 gene neighbour 法，成功地建立了酵母 SIR（silencing information regulator）作用网络和酵母 Prion 功能连锁网络[85]。

存在的问题

方法学上，二维凝胶电泳-质谱仍是目前最流行和较可靠的技术平台，但其通量、灵敏度和规模化均有待进一步加强。二维凝胶电泳有分离容量的先天限制，染色转移等环节操作困难费时，低丰度蛋白难以辨别，和质谱技术的联用已成为瓶颈。因此国际上开始重视研究以色谱/电泳-质谱为主的技术平台。另一方面，酵母双杂交技术虽已被用于研究蛋白质连锁群和蛋白质功能网络系统，但仍缺乏快速、高效的手段获取复杂蛋白质相互作用的多维信息。蛋白质组的生物信息学研究，虽然已有成功的先例，但其应用范围与准确率仍需提高，所面临的更大挑战是如何进行信息综合，准确分析蛋白质的相互作用，界定相互作用连锁群。

学术上，在基因组、转录组基础上的蛋白质组全谱研究，微生物已有成功的报道，但是在高等生物尤其是哺乳动物中未见报道，人类组织或细胞的蛋白质组全谱研究则基本未涉及。而由于物种演化中进化上的差别，人类基因组、转录组、蛋白质组的全景式比较对于不同层次上人类基因表达调控规律的认识必不可少；此外，人类基因组草图虽已公布，但是所估计的 3.5 万左右基因中一半以上纯属理论推测，需要从蛋白质组水平予以检验与确认，因此开展人类组织或细胞的蛋白质组表达谱的分析势在必行。

受基因调控的细胞内各种信号转导途径之间是相互交错和彼此关联的。虽然近年来人们对转导途径以及相互关系的认识取得了不少进展，但是针对任一生物体或组织/细胞开展全方位的蛋白质组相互作用网络的分析鲜有报道。而此类相互作用网络的揭示对于深刻认识重要生理、病理过程的机制不可缺少。

国内研究现状

在1995年国际上发表第一篇蛋白质组学的研究论文后不久，国家自然科学基金委即酝酿并于1997年设立了重大项目“蛋白质组学技术体系的建立”。在此前后，中国科学院生物化学研究所、军事医学科学院与湖南师范大学迅速启动了蛋白质组研究，建立并分别组合了二维电泳蛋白质组分离技术、2D-PAGE图像分析技术和蛋白质鉴定的质谱技术；先后举办了三次全国性的蛋白质组学术研讨会，并在国际上较早提出了功能蛋白质组学的研究战略。

经过几年努力，中国科学院上海生命科学研究院、军事医学科学院与复旦大学相继成立了专门的蛋白质组学研究中心。整体上，我国蛋白质组学技术平台的建设有了飞跃的发展，若干研究单位重点建立了技术平台，并在方法学的跟踪与创新上做了不少工作，使现有技术平台已经达到国际先进水平。我国科学家已经在蛋白质组分析技术与方法、在重大疾病如肝癌、维甲酸诱导白血病细胞凋亡启动模型及维甲酸定向诱导胚胎干细胞向神经系统分化的模型等的比较蛋白质组研究以及一些重要生理和病理体系的蛋白质组成分研究方面获得了不少成就，并在国际蛋白质组学核心刊物上发表了系列高水平的论文[86-92]。对不断涌现的新技术，如衍生出的多种分离和鉴定模式，我国已经进行了很好的跟踪和发展，在某些方面孕育着新的突破[93-105]。

我国的“疾病基因组学”研究已取得明显的成就，在神经系统遗传病致病基因、肝癌、心血管疾病、白血病等相关基因、造血干/祖细胞、下丘脑-垂体-肾上腺轴系统、海马体、胎肝、心血管和神经系统等组织或细胞的基因表达谱（即转录组）方面均做出了与国际前沿水平相当的工作[106-117]，且有我国的特色与优势。所取得的丰富数据将直接成为本项目开展对应蛋白质组学研究的基础与出发点。我国在重大疾病的功能蛋白质组学研究方面也取到了良好的起步，已进行肝癌细胞系及正常肝细胞蛋白质组的比较分析研究，发现了两者间不同的蛋白表达群[90,91]。此外，还进行了我国自行建立的肝癌高/低转移细胞系、肺癌高/低转移细胞系、原位食管癌/转移食管癌间的比较蛋白质组研究，初步发现了一批与肿瘤转移相关的蛋白质群。同时，心肌细胞与应激损伤的心肌细胞的比较蛋白质组研究也已有一定的基础。上述研究一方面证明我国已有的蛋白质组学技术平台已能支撑一定规模的研究，另一方面亦为我国在国际蛋白质组学研究领域争得了一席之地，同时为未来的发展奠定了良好的基础。

五、蛋白质组研究的发展趋势及我国的应对策略

蛋白质组研究的整体状况：相关技术已基本配套且基本达到实用化水平；已形成一定规模的专业队伍及专业机构，因此，其草创阶段已结束，发展阶段已开始；蛋白质组研究已成为后基因组时代最重大的生命科学命题之一；蛋白质组分

析已作为专门的技术体系广泛用于生命科学众多领域尤其是热点领域的研究；蛋白质组的基础研究与分析应用正以指数增长方式发展，其对现代生命科学的介入与贡献将可能使生命科学工作者从核酸时代逐渐回归蛋白质时代，使对生命系统与活动的分子机制的认识由间接的基因、核酸层次深入到生命的直接执行体——蛋白质层次，更将蛋白质研究无论是规模还是深度均推进到前所未有的程度。人们不难看到，它是基因组计划由结构走向功能的必然与必需，是生命科学由分析走向综合的必经之路，是连接微观分子系统、运行机制与宏观生物系统、生命活动的桥梁。因此，人们不难预期，蛋白质组研究将成为21世纪生命科学的重要支柱之一，其发展未可限量。

人们在憧憬蛋白质组研究美好未来的同时，也应看到“蛋白质组”作为“新生”领域在许多方面还很稚嫩，如2-DE的灵敏度虽然已达飞摩（fmol）级水平，但仍难将细胞组织内多种痕量调控蛋白分离显示出来，而此类蛋白对于基础与应用研究都极为重要（甚至比高含量结构蛋白更为重要），此方面仍需进一步改进；此外，现有质谱技术虽然在蛋白质组成分的鉴定中高效、灵敏、特异，但所用仪器价格十分昂贵，十倍于DNA自动测序仪，国内外只有极少数单位有能力购置，因而其推广受到很大的限制，此方面技术与仪器如不改善，将使蛋白质组研究与进展神速的基因组计划严重脱节；此外，蛋白质组研究不能局限于对已完成基因组计划的理论预测的蛋白质组进行实证分析，还应开辟“战场”，一方面对未完成或根本未进行基因组计划的重要生物进行前瞻性蛋白质组研究，以推动其基因组研究，另一方面大力加强重要生命活动中比较蛋白质组研究和重要组织的蛋白质组研究，后一方面将会有更大作为。

蛋白质组研究在国际上正如火如荼、轰轰烈烈，我国则刚启动。鉴于蛋白质组研究代表着生命科学研究中一个新的“制高点”，我国目前的状况应引起我们的高度警觉。虽然国家自然科学基金委已将“蛋白质组研究”作为重大项目立项，国家科技部亦将其列为国家基础研究重大项目，但相对国际上的大力突入与强大的竞争而言，这些只能是“九牛一毛”。21世纪生命科学在整个自然科学中的主导作用，我们注定要面对；蛋白质组研究在后基因组时代的支柱作用我们也必定要面对。我们与其到时“亡羊补牢”，不如现在就奋起直追。为此，综合考虑国际基因组与蛋白质组研究的现状与趋势以及我国基因组研究与蛋白质组研究相关技术的基础，作者认为我国的蛋白组研究不能重复或追随国际已有的工作，而应与我国现行的基因组研究及其他有我国特色或优势的生命科学研究紧密结合（但决不能仅仅融入这两大方面的项目中），走出自己的路。因此，建议开展如下工作：

1）*重大生命活动中蛋白质组的比较研究*　选取1～2类重大生命活动（如重大疾病）中几个相继的重要阶段，分别进行蛋白质组作图，进行系统的定性、定量比较，进而对差别蛋白进行鉴定，然后从核酸、蛋白两层

次对差别蛋白进行性能分析，从而确定重大生命活动的蛋白质组基础。

2）1～2种组织或细胞蛋白质组的系统研究　选择1～2种具有重要生物学意义或性状、且我国已完成cDNA大规模测序的组织或细胞，制备其高分辨率蛋白质组图并建立其蛋白质组图数据库，进而联合应用其cDNA大规模测序的数据，规模化研究其蛋白质组成。

3）蛋白质组的生物信息学研究　蛋白质组成员的序列、结构、功能及定位分类；基于生化途径、遗传网络等，构建蛋白质组功能系统即蛋白连锁图；建立人或其他哺乳动物蛋白质组数据库；高等生物基因组中蛋白质编码基因的识别及算法研究；基因翻译产物的结构、功能预测；基于蛋白质数据库与知识库的知识与规律发现。

4）蛋白质组分析的支撑技术研究　新型蛋白质结构、功能预测方法及程序；大规模蛋白质相互作用分析技术；蛋白质分析鉴定中新型质谱技术的发展及应用；基于蛋白质序列、结构及蛋白质组数据库的知识与规律发现理论与方法；HTS（highthroughout system）系统及蛋白质组分析自动操作系统（蛋白质组分析机器人）等。

总之，我国在蛋白质化学与蛋白质科学（protein science）领域曾取得举世瞩目的成就，近年在蛋白质组学这一新兴领域已有了良好的基础，一批研究结果已在国际蛋白组学的核心期刊发表，技术平台的部分指标达到国际先进水平。但是，现有的规模与层次难以提供我国医药卫生事业与生物技术产业迅猛发展所急需的、强有力的蛋白质组学学术与技术支撑，难以适应我国基因组学等现代生命科学前沿领域对蛋白质组学的广泛需求，难以应对国际在蛋白质组学这一战略“高地”的激烈竞争。因此，国家层面上对蛋白质组学的部署不容迟缓。

参考文献

1 Wasinger V C, Humphery Smith I, Williams K L, *et al*. Progress with gene product mapping of the molicutes: *Mycoplasma genitalium*. Electrophoresis, 1995, 16: 1090～1094

2 Kahn P. From genome to proteome: Looking at a cell's proteins. Science, 1995, 270: 369～370

3 Swinbanks D. Government backs proteome proposal. Nature, 1995, 386: 653

4 Nowak R. Entering the postgenome era. Science, 1995, 270

5 Chait B T. Trawling for proteins in the post-genome era. Nature Biotech, 1996, 14: 1544

6 Murray J C, Buetow K H, Weber J L, *et al*. A comprehensive human map with centimorgan density. Science, 1994, 256: 2049～2054

7 Hudson T J, Stein L D, Gerety S S, *et al*. An STS based map of the human genome. Science, 1995, 270: 1945～1954

8 Chumakow I M, Rigoult P, Le Gall I, *et al*. A YAC contig map of the human genome. Nature, 1995, 377 (Suppl): 175～297

9 Venter J V, Smith H, Hood L. A new strategy for genome sequencing. Nature, 1996, 318: 364～366

10 Fleischmann R D, Adams M D, White O, *et al*. Whole genome random sequencing and assembly of *Haemophilis influenzae* Rd. Science, 1995, 269: 496～512

11 Fraser C M, Gocayne J K, Whit O, *et al*. The minimal gene complement of Mycoplasma genitalium. Science, 1995, 270: 297～403

12 Blattmer F B, G Plunkett Ⅲ, Bloch C A, *et al*. The complete genome sequence of E. coli K 12. Science, 1997, 277: 1453～1462

13 Goffeau, A, Barrel B G, Bussey H, *et al*. Life with 6000 genes. Science, 1996, 274: 546～567

14 Mewes H W, Albermann K, Bahr M, *et al*. Overview of the yeast genome. Nature, 1997, 387 (suppl): 7～65

15 Wilson R, Ainscough R, Anderson K, *et al*. 2.2 Mb of contiguous nucleotide sequence from chromosome Ⅲ of *C. elegans*. Nature, 1994, 368: 32～38

16 Wodicka L, Dong H, Mittmann M, *et al*. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. Nature Biotech, 1997, 15: 1359～1367

17 Ramsay G. DNA chips: State-of-the-art. Nature Botech, 1998, 16: 40～44

18 Velculescu F B, Zhang L, Vogelstein B, *et al*. Serial analysis of gene expression. Science, 1995, 270: 484～4872

19 Perler F B, Olsen G J, Adam E. Compilation and analysis of intein sequences. Nucleic Acid Res, 1997, 25: 1087～1093

20 Alberts B. The cell as a collection of protein machines: Preparing the next generation of molecular biologists. Cell, 1998, 92: 291～294

21 O'Farrell P H. High resolution two dimensional electrophoresis of proteins. J Biol Chem, 1975, 250: 4007～4021

22 Strupat K, Karas M, Hillenkamp F, *et al*. Matrix-assisted laser desorption ionization mass spectrometry of proteins electroslotted after PAGE. Anal Chem, 1994, 66: 464～470

23 Mann, M, Hojrup P, Roepstorff P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biol Mass Spectrometry, 1993, 22: 238～245

24 Wilkins M R, Sanchez J C, Gooley A A, *et al*. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotech & Genetic Engineering Rev, 1995, 13: 19～50

25 Geisow M. Proteomics. One small step for a digital computer, one giant leap for humankind. Nature Biotech, 1998, 16: 206

26 Humphery Smith I, Cordwell S J, Blackstock W P. Proteome research: Complementarity and limitations with respect to the RNA and DNA worlds. Electrophoresis, 1997, 18: 1217～1242

27 Shevchenko A, Mann M, Boucherie H, *et al*. Linking genome and proteome by mass spectrometry: large scale identification of yeast proteins from two dimensional gels. PNAS, 1996, 93: 14440～14445

28 Celis J E, Rasmussen H H, Vanderckhove J, *et al*. Human 2-D PAGE databases for proteome analysis in health and disease: http://biobose. dik/egibin/celis. FEBS Lett, 1996, 398: 129～134

29 Bin L, Zwilling R, Pallin V, *et al*. Two dimensional gel electrophoresis of *Caenorhabditis elegans* homogenates and identification of protein spots by microsequencing. Electrophoresis, 1997, 18: 557～562

30 Langen H, Fountoulakis M, Takacs B, *et al*. From genome to proteome: protein map of *Haemophinlus influenzae*. Electrophoresis, 1997, 18: 1184～1192

31 Cordwell S J, Humphery Smith I, Shaw D C, *et al*. Characterization of basic proteins from *spiiroplasma melliforum* using novel immobilized pH gradients. Electrophoresis, 1997, 18: 1393～1398

32 Yan J X, Williams K L, Hochstrasser O F, *et al*. The *Dictyostelium discoidenm* proteome the SWISS-2D

PAGE database of the multicellular aggregate (Slug). Electrophoresis, 1997, 18: 491～497

33 O'Connor C D, Qi S Y, Fowler R, *et al*. The proteome of *Salmonella* enteria serovar typhimurium: current progress on its determination and some applications. Electrophoresis, 1997, 18: 1483～1490

34 Guerreiro N, Djordievic M A, Rolfe B G, *et al*. New *Rhizobium leguminosarum* flavonoid induced proteins revealed by proteome analysis of differentially displayad proteins. Mol Plant Microbe Interact, 1997, 10: 506～516

35 Link A J, Yates J R 3rd, Carmack E B, *et al*. Identifying the major proteome components of *Haemophilus influenzae* type strain NCTC 8143. Electrophoresis, 1997, 18: 1314～1334

36 Link A J, Church G M, Robison K. Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. Electrophoresis, 1997, 18: 1259～1313

37 Garrels J I, Payne W E, Mesquita Fuentes R, *et al*. Proteome studies of *Saccharomyces cerevisiae*: identification and characterization of abundant proteins. Electrophoresis, 1997, 18: 1347～1360

38 Sazuka T, Ohara O. Towards a proteome project of *cyanobacteriium Syoechocysitis* sp strain PCC6803: linking 130 protein spots with their respective genes. Electrophoresis, 1997, 18: 1252～1258

39 Van Bogelen R A, Neidhardt F C, Olson E R, *et al*. *Escherichia coli* proteome analysis using the gene-protein database. Electrophoresis, 1997, 18: 1243～1251

40 Dainese P, James P, Kertesz M, *et al*. Probing protein function using a combination of gene knockout and proteome analysis by mass spectrometry. Electrophoresis, 1997, 18: 832…842

41 Blomberg A. Osmoresponsive proteins and functional assessment strategies in *Saccharomyces cerevisiae*. Electrophoresis, 1997, 18: 429～440

42 Ostergaard M, Celis J E, Wolf H, *et al*. Proteome profiling of bladder squamous cell carcinomas: identification of markers that define their degree of differentiation. Cancer Res, 1997, 57: 4111～4117

43 Wimmer K, Hanash S M, Thoraval D, *et al*. Two-dimensional separations of the genome and proteome of neuroblastoma cells. Electrophoresis, 1996, 17: 1741～1751

44 Urquhart B L, Haumphery-Smith I, Brithon W L, *et al*. "Proteomic contig" of *Mycobacterium tuberculosis* and *mycobacterium bovis* (bcg) using novel immobilized pH gradients. Electrophoresis, 1997, 18: 1384～1392

45 Corduell S J, Humphery-Smith I, Basseal D J. Proteome analysis of *Spiroplasma melliferum* (A56) and protein characterization across species boundaries. Electrophoresis, 1997, 18: 1335～1346

46 Cordwell S J, Humphery-Smith I. Evaluation of algorithms used for cross-species proteome characterization. Electrophoresis, 1997, 18: 1410～1417

47 Wasinger V C, Humphery-Smith I, Bjellqist B. "Proteomic contig" of *Ochrobactrum anthropi*, application of extensive pH gradients. Electrophoresis, 1997, 18: 1373～1383

48 Payne W E, Garrels J I. Yeast protein database (YPD): a database for the complete proteome of *Saccharomyces cerevisiae*. Nucleic Acids Res, 1997, 25: 57～62

49 Fromont-Racine M, Legrain P, Rain J C. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nature Genetics, 1997, 16: 277～282

50 Lecenier N, Foury F, Goffeau A. Two-hybrid systematic screening of the yeast proteome. Bio-Essays, 1996, 20: 1～6

51 Packer N H, Wiilliams K L, Redmond J W, *et al*. Proteome analysis of glycoforms: a review of strategies for the microcharacterisation of glycoproteins separated by two-dimensional PAGE. Electrophoresis, 1997, 18: 452～460

52 Qi S Y, O'Connor C D, Moir A. Proteome of *Salmonella typhimurium* SL1344: identification of novel

abundant cell envelope proteins and assignment to a two-dimensional reference map. J Bactoriol, 1996, 178: 5032～5038

53 J X, Hochstrasser D F, Pasqual C, *et al*. Large scale amino-acid analysis for proteome studies. J Chromatogr, 1996, 736: 291～302

54 Abbott A. And now for the proteome. Nature, 2001, 409: 747

55 Fields S. Proteomics in genotneland. Science, 2001, 291: 1221

56 Yuen Ho, *et al*. Systematic identification of *protein complexes in Saccharomyces cerevisiae* by mass spectrometry. Nature, 2002, 415: 180～183

57 Akhilesh Pandey, Matthias Man. Proteomics to study genes and genomes. Nature, 2000, 405: 837～846

58 Potter Wickware, Paul Smaglik. Proteomics to study genes and genomes. Nature, 2001, 413: 869～875

59 Anton J, Ioannis Iliopoulos, Nikos C K, Christos A O. Protein interaction maps for complete genomes based on gene fusion events. Nature, 1999, 402: 86～90

60 Anne-Claude Gavin, *et al*. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature, 2002, 415: 141～147

61 Service, Robert F. PROTEOMICS: A Proteomics Upstart Tries to Outrun the Competition. Science, 2001, 294: 2079～2080

62 Service, Robert F. PROTEOMICS: *Proteomics 2.0*: The View Ahead. Science, 2001, 294: 2076

63 Ideker, Trey, Thorsson, Vesteinn, Ranish, Jeffrey A., Christmas, Rowan, Buhler, Jeremy, Eng, Jimmy K., Bumgarner, Roger, Goodlett, David R., Aebersold, Ruedi, Hood, Leroy. Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network. Science, 2001, 292: 929～934

64 Gerstein, Mark, Lan, Ning, Jansen, Ronald. PROTEOMICS: Enhanced: Integrating Interactions. Science, 2002, 295: 284～287

65 Marshall, Eliot. PROTEOMICS: A Physicist-Turned-Biologist. Science 2001, 294: 2085

66 Service, Robert F. PROTEOMICS: Gene and Protein Patents Get Ready to Go Head to Head. Science, 2001, 294: 2082～2083

67 Service, Robert F. PROTEOMICS: Searching for Recipes for Protein Chips. Science, 2001, 294: 2080～2082

68 Service, Robert F. PROTEOMICS: High-Speed Biologists Search for Gold in Proteins. Science, 2001, 294: 2074～2077

69 Mirène Fauchon, Gilles Lagniel, Jean-Christophe Aude, Luis Lombardia, Pascal Soularue, Cyrille Petat, Gérard Marguerie, André Sentenac, Michel Werner, and Jean Labarre. Sulfur Sparing in the Yeast Proteome in Response to Sulfur Demand. Molecular Cell, 2002, 9: 713～723

70 Mary B D, Robert F O. The search for predictive patterns in ovarian cancer: Proteomics meets bioinformatics. Cancer Cell, 2002, 1: 111～112

71 Archa H. Fox, Yun Wah Lam, Anthony K. L. Leung, Carol E. Lyon, Jens Andersen, Matthias Mann, and Angus I. Lamond. Paraspeckles: A Novel Nuclear Domain. Current Biology, 2002, 12: 13～25

72 Jens S. Andersen, Carol E. Lyon, Archa H. Fox, Anthony K. L. Leung, Yun Wah Lam, Hanno Steen, Matthias Mann, and Angus I. Lamond. Directed Proteomic Analysis of the Human Nucleolus. Current Biology, 2002, 12: 1～11

73 Jason K. Ospina and A. Gregory Matera. Proteomics: The Nucleolus Weighs In. Current Biology, 2002, 12: R29～R31

74 Miroslav Dundr and Tom Misteli. Nucleolomics: An Inventory of the Nucleolus. Molecular Cell, 2002, 9: 5～7

75 Piyanun Harnpicharnchai, Jelena Jakovljevic, Edward Horsey, Tiffany Miles, Judibelle Roman, Michael Rout, Denise Meagher, Brian Imai, Yurong Guo, Cynthia J. Brame, Jeffrey Shabanowitz, Donald F. Hunt, and John L. Woolford Jr. Composition and Functional Characterization of Yeast 66S Ribosome Assembly Intermediates. Molecular Cell, 2001, 8: 505～515

76 Mark R. Bray, Stuart Bisland, Subodini Perampalam, Wai-May Lim, and Jean Gariépy. Probing the surface of eukaryotic cells using combinatorial toxin libraries. Current Biology, 2001, 11: 697～701

77 Timothy S. Lewis, John B. Hunt, Lauren D. Aveline, Karen R. Jonscher, Donna F. Louie, Jennifer M. Yeh, Theresa S. Nahreini, Katheryn A. Resing, and Natalie G. Ahn. Identification of Novel MAP Kinase Pathway Signaling Targets by Functional Proteomics and Mass Spectrometry. Molecular Cell, 2000, 6: 1343～1354

78 Robert F. Proteomics, can Celera do it again? Science, 2000, 287: 2136

79 Haynes PA. Proteome profiling-pitfalls and progress. Yeast, 2000, 17: 81

80 Seow T K, *et al*. Two-dimensional electrophoresis map of the human hepatocellular carcinoma cell line, HCC-M, and identification of the separated proteins by mass spectrometry. Electrophoresis 2000, 21: 1787

81 Bakhtiar R, *et al*. ESI-MS and MALDI-MS——Emerging technologies in biomedical science. Biochem Pharmacol, 2000, 8: 891

82 Szallasi Z. Gene expression patterns and cancer. Nature Biotech, 1998, 16: 1292

83 Albertha J M, *et al*. Protein interaction mapping in *C. elegens* using protins involved in vulval development. Science, 2000, 287: 116

84 Houry W A, *et al*. Identification of in vivo substrates of the chaperonin. Nature, 1999, 402: 147

85 Eisenberg D, *et al*. Protein function in the post-genomic era. Nature, 2000, 405: 823

86 Xia Q C, *et al*., Application of free-flow electrophoresis to the purification of trichosanthin from a crude product of acetone fractional precipitation. Electrophoresis, 1998, 19: 1097

87 Xia Q C, *et al*. Determination of sialic acid content in glycoproteins by capillary zone electrophoresis. Electrophoresis, 1999, 20: 2930

88 Wan J H, *et al*. Proteomic analysis of apoptosis initiation induced by all-trans retinoic acid in human acute promyelocytic leukemia cells. Electrophoresis, 2001, 21: 3026

89 Guo X X, *et al*. Proteomic Characterization of Early Stage Differentiation of Mouse Embryonic stem Cells into Neural Cells Induced by Retinoic Acid in vitro. Electrophoresis, 2001, 21: 3067

90 Yu L R, Xia Q C, *et al*. Identification of differentially expressed proteins between human hepatoma and normal liver cell lines by two-dimensional electrophoresis and liquid chromatography-ion trap mass spectrometry. Electrophoresis, 2000, 21: 3058

91 Yu L R, Xia Q C, *et al*. Proteome altration in human hepatoma cells transfected with antisense EGF receptor sequence. Electrophoresis, 2000, 22: 3001

92 Zou H F, Zhang Y K, *et al*. Quantitative Study of Competitive Binding of Drugs to Protein by Microdialysis/HPLC. J Chromatogrphy A, 1999, 849: 599～608

93 Zou H F, Zhang Y K, *et al*. Comparative study on the distribution of ovalbumin glycoforms by capillary electrophoresis. Anal Chem, 1998, 70: 373

94 Zou H F, Zhang Y K, *et al*. Study of Physically Adsorbed Stationary Phase for Open-Tubular Capillary Electrochromatography. Electrophoresis, 1999, 20: 2891

95 Liang S P, *et al*. Identification of venom proteins of spider Shuwena on two-dimensional electrophoresis gel by N-terminal microsequencing andmass spectrometric peptide mapping. J Protein Chem, 2000, 3: 225～230

96 Liang S P, *et al*. Analysis of recombinant and modified proteins by capillary zone electrophoresis/electrospray ionization-tradem mass spectrometry. J Chomatography A, 1999, 855: 695

97 Zou H F, Zhang Y K, *et al*. Capillary Electrochromatography Using a Strong Cation-Exchange Column with a Dynamically Modified Cationic Surfactant. Anal Chem, 2000, 72: 616

98 Zou H F, Zhang Y K, *et al*. On-Column Enrichment In Capillary Electrochromatography. Anal Chem, 2000, 72: 23

99 Yang P Y, *et al*. Pulsed-electrospray mass spectrometry. Anal Chem, 2001, 73: 4748

100 Zhang X M, *et al*. Comprehensive Two-Dimensional Capillary LC and CE for Resolution of Neutral Components in Traditional Chinese Medicines. J Sep Sci, 2001, 24: 10

101 Zhang X M, *et al*. Single Step On-Column Frit Making For Capillary High Performance Liquid Chromatography Using Sol-Gel Technology. J Chromatogr A, 2001, 910: 13

102 Kong J L, *et al*. Direct Electrochemistry of Co-factor Redox Sites in a Bacterial Photosynthetic Reaction Center Protein. J Am Chem Soc, 1998, 120: 7371

103 Huo K K, *et al*. Two DNA replication origin identified in the K1 killer toxin gama subunit gene promoter. Current Genetics, 1999, 35: 335

104 Huo K K, *et al*. Protein disulfide isomerase genes of *Kluyveromyces lactis*. Yeast, 2000, 16: 329

105 Hu R M, Han Z G, Song H D, *et al*. Gene expression profiling in the human hypothalamus-pituitary-adrenal axis and full-length cDNA cloning. Proc Natl Acad Sci USA, 2000, 97: 9543～9548

106 Yu Y T, Zhang C G, Zhou G Q, Wu S F, Qu X H, Wei H D, Xing G C, Zhai Y, Wan J H, Ouyang S G, Li L, Zhang S W, Wu C T, He F C. Gene Expression Profiling in Human Fetal Liver and Identification of Tissue- and Developmental Stage-specific Genes through Compiled Expression Profiles and Efficient Cloning of Full-Length cDNAs. Genome Res, 2001, 11 (8): 1392

107 Mao M, *et al*. Identification of genes expressed in human CD34 + hematopoietic stem/protgenitor cells by expressed sequence tags and efficient full-length cDNA cloning. Proc Natl Acad Sci USA, 1998, 95: 8175

108 Huo K K *et al*. Isolation of a novel candidate oncogene within a frequently amplified region at 3q26 in ovarian cancer. Cancer Research, 2001, 61: 3806

109 Li Y, *et al*. Stimulation of the mitogen-activated protein kinase cascade and tyrosine phosphorylation of the epidermal growth factor receptor by hepatopoietin. J Biol Chem, 2000, 275: 37443

110 Wang G, *et al*. Identification and characterization of receptor for the mammalian hepatopoietin that is homologous to yeast ERV1. J Biol Chem, 1999, 274: 11469

111 Liu X Q, *et al*. Characterization of ARF GAP family and its first member derived from human being. FEBS LETT, 2001, 490: 79～83

112Zhang C G, *et al*. Characterization, Chromosomal Assignment, and Tissue Expression of a Novel Human Gene Belong to ARF GAP Family. Genomics, 2000, 63 (3): 400

113 Qu X H, *et al*. Characterization and Tissue Expression of a Novel Human Orthologue of Mouse npdc1. Gene, 2001, 264: 37

114 Zhu F X, *et al*. cDNA transfection of amino-terminal fragment of urokinase efficiently inhibits cancer cell invasion and metastasis. DNA & Cell Biology, 2001, 20: 297

115 Luo T H, *et al*. A genome-wide search for type Ⅱ diabetes susceptibility genes in Chinese Hans. Diabetologia, 2001, 44: 501

116 Tang J G, *et al*. Anti-autolysis of trypsin by modification of autolytic site Arg117. Biochem Biophys Res Commun, 1998, 250: 235

117 Tang J G, *et al*. Separation and characterization of trypsin and carboxypeptidase B digested products of Met-Lys-human proinsulin. Applied Biochemistry and Biotechnology, 1999, 76: 107

（贺福初）

第二章　蛋白质组研究方法

第一节　概　　述

蛋白质组的研究远比基因组研究复杂。一方面，蛋白质的数目远大于基因的数目，这是由于基因的拼接和翻译后的修饰造成的。人的基因组有 25 000～40 000个编码基因[1]，其表达的蛋白质可以达十几万甚至更多。另一方面，基因是相对静态的，一种生物体仅有一个基因组，而蛋白质是动态的，随时间、空间的变化而变化。一个细胞中的蛋白质可多达上万种，而它们的拷贝数可能相差几百倍到几十万倍。组成蛋白质的氨基酸有 20 种，加上修饰的氨基酸就更多，而 DNA 仅由 4 种核苷酸组成。从技术手段上讲，人们无法采用在基因研究中所普遍采用的 PCR 技术来使微量的蛋白质得到扩增，至今尚没有一种蛋白质的测序技术可与在基因组研究中起关键作用的自动化的 DNA 测序技术相媲美。DNA 微阵列技术的发展与应用，使得基因的筛选实现了高通量，而目前的蛋白质分析技术离工业规模的高通量还有较大差距（表 2-1）。

所幸的是，尽管核酸研究的技术近年来取得了突飞猛进的发展，蛋白质的研究技术并未停滞不前。由 O'Farrel 等在 1975 年建立的双向电泳技术（two dimensional electrophoresis，2-DE）可同时分离数千种蛋白[2]，20 世纪 80 年代固相化 pH 梯度凝胶的引进使得双向电泳的重复性和加样量得到巨大的改善，从而使今天的蛋白质组研究得以实施[3]。以计算机技术为基础的多种图像分析与大规模数据处理软件的问世，使得科学家们处理复杂的类似“满天星”样的蛋白质图谱并建立相应的数据库得心应手。80 年代后期出现的两种软电离质谱技术-基质辅助激光解吸电离飞行时间质谱（matrix assisted laser desorption ionization time of flight mass spectrometry，MALDI-TOF-MS）[4]与电喷雾电离质谱（electro-spray ionization mass spectrometry，ESI-MS）[5]，可精确测定生物大分子的相对分子质量及多肽序列，使得微量快速的蛋白质鉴定得以实现。回想 10 年以前，一个蛋白质化学家一年鉴定 2～3 个蛋白质，而现在，蛋白质组研究技术加上基因组提供的信息使得一个科学家一个星期可以鉴定几百个蛋白质。

双向电泳技术、计算机图像分析与大规模数据处理技术以及质谱技术被称为蛋白质组研究的三大基本支撑技术。蛋白质组研究的技术路线如图 2-1。从细胞、体液或组织等生物样品中提取的蛋白质，经 2-DE 分离，染色，得到蛋白质表达谱。采用计算机图像分析技术，可对图谱上的蛋白质点进行定位、定量、图谱比

表 2-1　DNA 分析（cDNA 微阵列）与蛋白质分析（2-DE 与质谱）方法的比较[6]

	DNA	蛋白质
检测水平/灵敏度	DNA 可用 PCR 技术扩增 检测限约为 30 拷贝/样本（transcripts）	蛋白质无法被扩增
检测方法	高度特异 DNA 与其互补序列（DNA/RNA）的杂交能力使得 DNA 很容易固相化在靶上并用荧光探针检测。高密度阵列检测专用的扫描仪已经出现	非特异 蛋白质可以用非特异的方法如蛋白染色或较特异的方法如抗体进行检测
分子的稳定性	DNA 是非常稳定的，即使在提高温度的情况下也不丢失生物活性	蛋白质在变性情况下将丧失天然功能。蛋白质需要格外小心以保持其天然构象
重复性	重复性好 将DNA 固相化在表面已经是成熟的技术，加之 DNA 固有的稳定性，方法的重复性较好	重复性不好 由于蛋白质分子的稳定性不好，方法的重复性需格外关注
消耗	初始的投入是较大的（微阵列、扫描仪、共聚焦显微镜），但短期内得到的数据量相当大	目前的蛋白质分析方法依赖昂贵的仪器设备（质谱仪、2-DE 装置）
专业化	方法已相当标准化	蛋白质的分析较困难。大部分分析工作如纯化与质谱分析等都需要受过训练的实验室人员来操作
时间/自动化	已成为高通量的扫描技术	目前还不能达到工业需求的高通量
翻译后修饰	无法研究	研究翻译后修饰的惟一方法是研究蛋白质本身
动态范围	5 个数量级	7～12 个数量级

较、差异点寻找等。对胶上蛋白质点的鉴定可以采用两种不同的技术路线。一种是采用膜转印技术，将电泳胶上的蛋白质转印到膜上（硝酸纤维素膜或 PVDF 膜），然后采用 Edman 降解或氨基酸组成分析等传统生物化学的方法进行分析，将分析结果输入特定的程序并进行数据库检索，从而实现对蛋白质的鉴定。另一种是以质谱为基本技术的蛋白质鉴定技术路线。将胶上蛋白质点直接进行蛋白酶切，回收酶切肽段，采用 MALDI-TOF-MS 测定肽质量指纹图（peptide mapping fingerprint，PMF），或者采用电喷雾串联质谱（ESI-MS-MS）测定肽序列，分别进行数据库检索，实现蛋白质的鉴定。如果是蛋白质数据库中不存在的蛋白质，还可以通过特定的检索程序进行基因数据库的分析检索。以质谱技术为基础

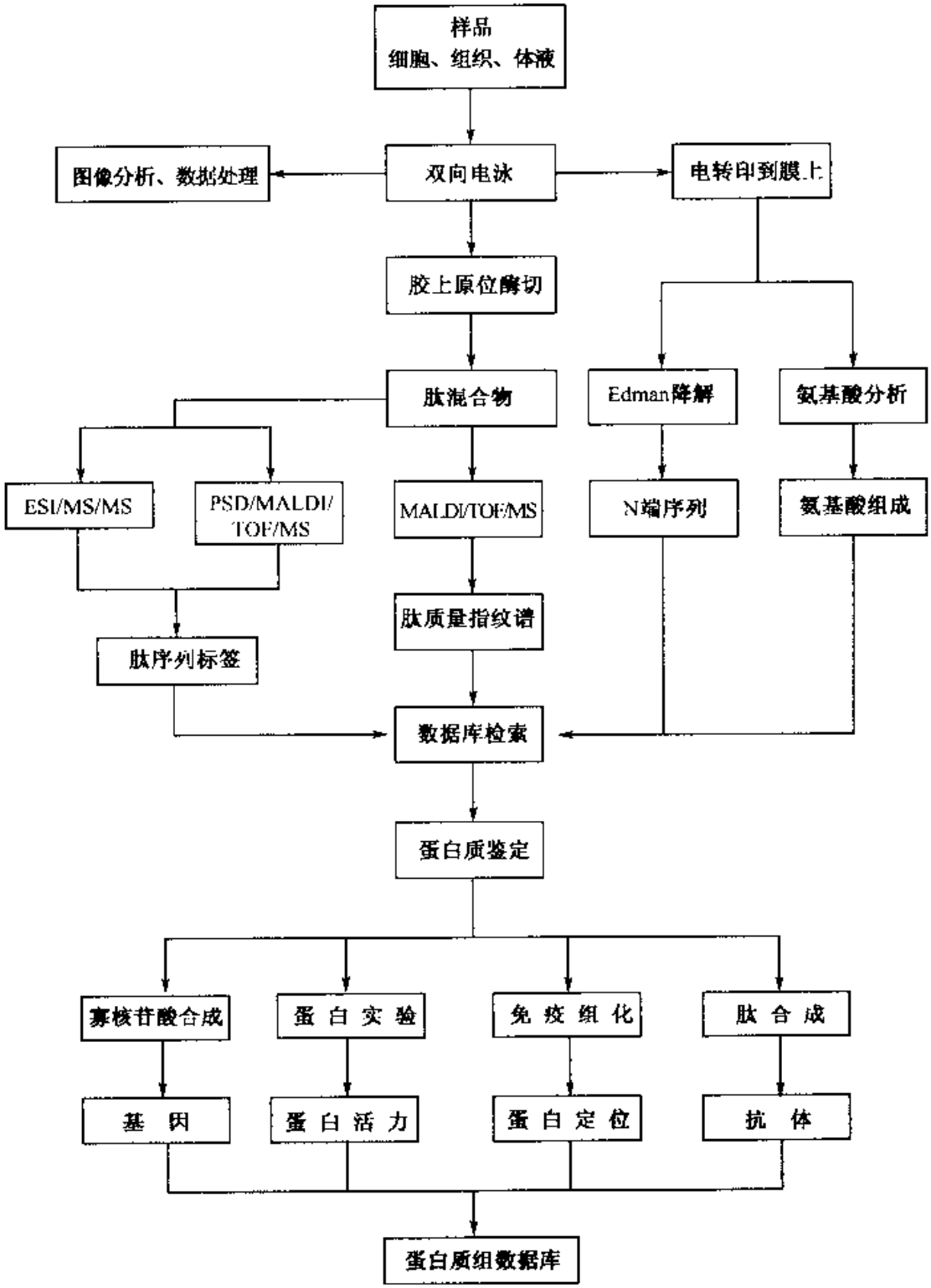


图 2-1　蛋白质组研究的技术路线流程图

的蛋白质鉴定技术，由于其灵敏度高（可以达到皮摩尔到飞摩尔，10^{-12}～10^{-15}），速度快，易实现自动化，已经成为蛋白质组研究中主要的蛋白质鉴定技术。对已鉴定的蛋白质，还要采用生物学或生物化学的方法进行进一步功能验证。

除了上述蛋白质组研究的基本技术外，新的蛋白质组研究技术也不断出现，如用于定量蛋白质组学研究的同位素编码亲和标签技术（isotope-coded affinity tags, ICAT）[7]和双色荧光技术[8]等；用于蛋白质-蛋白质相互作用研究的酵母双杂交技术[9]、蛋白质复合物免疫分离与质谱鉴定技术[10]；用于大规模蛋白质分离与鉴定的多维色谱-质谱联用技术[11]；用于翻译后修饰如磷酸化、糖基化蛋白质图谱展示与检测技术[12,13]以及蛋白质芯片技术[14]等。蛋白质组研究的各项基本技术，本书后续各章将有详细的介绍，本章将侧重对蛋白质组研究现有技术的改进以及部分新技术方法加以介绍。

第二节　大规模的蛋白质分离技术

蛋白质组学的目的是实现对一个基因组所编码的全部蛋白质及其相互作用的研究。以简单的真核生物酵母为例，其基因组编码蛋白质为6145个[15]。人类基因组约含28 000～40 000个编码基因，其表达的蛋白质可能达十几万[16]。如何实现对细胞、组织或体液所含有的成千上万个蛋白质的有效分离，是蛋白质组研究所要解决的首要问题。理想的蛋白质分离方法首先要具备超高分辨率，能够将成千上万个蛋白质包括它们的修饰物同时分离并与后续的鉴定技术有效衔接。这种理想的分离方法还应当对不同类型的蛋白质，包括酸性的、碱性的、疏水的、亲水的，相对分子质量小的、相对分子质量大的均能有效分离。

一、二维凝胶电泳技术

二维凝胶电泳（two-dimensional gel electrophoresis, 2-DE）是目前惟一能将数千种蛋白质同时分离与展示的分离技术。2-DE是由O'Farrell和Klose于1975年分别提出的[2,17]。在2-DE中，蛋白质首先根据等电点的不同在一向等电聚焦电泳中分离，接着被转移到二向SDS-聚丙烯酰胺凝胶上，再根据相对分子质量大小不同被分离。由于电泳装置的限制和载体两性电解质的特性，使得2-DE的重复性很差，很难制备大量重复性好的2-DE胶。20世纪80年代初期出现的固相化pH梯度（immobilized pH gradient, IPG）等电聚焦电泳技术，使2-DE的重复性和上样量大大改善[3]，不同实验室之间的实验结果可以进行比较，加之后续的微量鉴定技术如Edman微量测序技术以及质谱技术灵敏度的提高，使得2-DE获得了广泛的应用，成为蛋白质组研究中首选的分离技术（详见第四章）。

尽管如此，2-DE 技术仍存在许多不足，如膜蛋白、碱性蛋白、低丰度蛋白的分离与检测、重复性以及规模化和自动化的问题。对 2-DE 技术的改进，包括对 2-DE样品的前处理及 2-DE 后的蛋白点检测的研究一直在进行。

1．低丰度蛋白、膜蛋白的检测与样品预分离技术

大部分蛋白质是低丰度蛋白质，而低丰度蛋白质往往是执行重要生物功能的蛋白质，如调控蛋白、信号转导蛋白和受体蛋白等。对酵母基因组和蛋白质组研究的结果显示，密码子偏性系数（codon bias index，CBI）小于 0.2 的基因编码的蛋白质为低丰度蛋白质，酵母基因组编码的全部 6000 多个蛋白质有一半的密码子偏性系数小于 0.2，其中 2500 个编码基因的 CBI 小于 0.1。这些蛋白质在传统的 2-DE 上是看不到的[18]。尽管使用窄范围 pH 胶条和加大上样量可适当提高低丰度蛋白的检出，但由于 2-DE 胶的上样量受到胶的负载量的限制，因而不可能无限制加大上样量。蛋白质的检出限与染色方法、起始加样量之间的关系见表 2-2。

表 2-2　蛋白质的检出限与染色方法、起始加样量的关系[17]

	银　染[a]		考马斯亮蓝染[a]	
蛋白质的丰度	蛋白质的量	细胞数	蛋白质的量	细胞数
拷贝数/细胞	mg[b]		mg[b]	
10	20.073	1.2×10^{9}	2007.3	1.2×10^{11}
100	2.007	1.2×10^{8}	200.7	1.2×10^{10}
1000	0.201	1.2×10^{7}	20.1	1.2×10^{9}
10 000	0.020	1.2×10^{6}	2.0	1.2×10^{8}
100 000	0.002	1.2×10^{5}	0.2	1.2×10^{7}

[a]银染蛋白质的检测限为 1 ng，考染蛋白质的检测限为 100 ng。

[b]以 6×10^{7} 个细胞得到 1 mg 可溶酵母蛋白计算。蛋白质的分子质量以 50 kDa 计算。

造成低丰度蛋白很难被检测到的原因主要是由于高丰度蛋白质的干扰和低丰度蛋白质的低溶解度。大部分基于 2-DE 的蛋白质组研究方法都采用“一步法(one-extract-one-gel)”，即一步提取的细胞总蛋白在一块胶上进行电泳分离。这种方法所能识别和检测到的大部分是高丰度的蛋白质。采用窄范围的 IPG 胶条(2～3 个 pH 单位）和非常窄范围的 IPG 胶条（约 1 个 pH 单位），增加了电泳胶的分辨率，加大了样品负载量，使得一些原先观察不到的蛋白质可以被观察到。但是由于样品中相当数量的蛋白质的等电点超出了胶条的 pH 范围，因而无法达到聚焦并沉淀在胶条的两端，从而大大影响了一向等电聚焦电泳的质量。另外仍需要将高丰度蛋白质去除，以避免高丰度蛋白质对低丰度蛋白检测的干扰。

膜蛋白是一些疏水性的蛋白质。当采用常规的 2-DE 方法进行分离时，由于膜蛋白在常规提取液中的低溶解度，另外因聚焦而产生浓缩、聚集，使膜蛋白很难溶解并进入二向 SDS-PAGE 电泳，因而很难被检测到。

对全细胞蛋白质采用预分离（pre-fractionation）是提高低丰度蛋白质和膜蛋白检出限的有效方法。一种是被称为“分步提取法（sequential extraction）”的蛋白质预分离技术，其原理是基于蛋白质的溶解度不同，采用 3 种不同的提取液对全细胞蛋白质进行分布提取和分别电泳（图 2-2）。采用分步提取法对细胞总蛋白进行预分离，实际上是在 2-DE 根据蛋白质等电点和分子质量不同进行二维分离的基础上，增加了溶解性能不同的第三维分离。我们的实验结果显示，采用三步法，观察到的蛋白质点增加了约 20%，很多低丰度的蛋白质可以被观察到，特别是一些疏水性的蛋白质[19]。

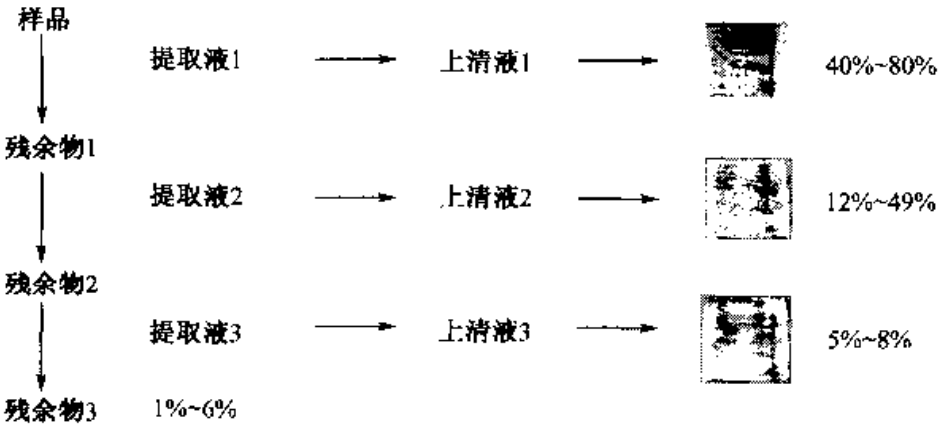


图 2-2　蛋白质分步提取-电泳法流程

另一种对蛋白质采用预分离的方法称为“多间隔电解法（multi-compartment-electrolyser）”[20]。此种方法实质上是一种制备等电聚焦的方法。电泳装置由中间的聚焦室和两端的电极室组成。聚焦室由隔膜分成多个样品室。电极室内装电极液，由阴阳离子交换膜将电解质与样品分开。操作时，将预分离的蛋白质混合物放进聚焦室中，随着等电聚焦电泳的进行，混合物中的蛋白质根据各自的等电点不同而聚焦到相应的样品室中并被分别收集。将经过此种预分离的蛋白样品采用窄范围 IPG 胶条进行 2-DE 分离，与未经预分离样品的 2-DE 谱相比，分辨率和胶的质量均明显改善。

2. 碱性蛋白质的分离与检测

碱性蛋白质在传统的 2-DE 中很难被分开。这主要是由于商业化的 IPG 胶条的 pH 范围通常在 3～10。等电点超过 10 的蛋白质很难被分开。这个问题由于 IPG 胶条的 pH 范围目前已经扩大到 12 而获得部分的解决。另一个影响碱性蛋

白质分离与检测的原因，是由于大部分细胞提取物是酸性蛋白质，而酸性蛋白质在碱性 pH 梯度胶上不能很好地溶解，特别是在上样量较大的情况下。可以通过将酸性蛋白质去除的办法改善碱性蛋白质的分离与检测。如采用上述提到的样品预分离的方法，如制备等电聚焦预分离，可将碱性蛋白质富集，再采用碱性 pH 梯度胶进行分离。

3．IPG 胶条的改进

为了提高 2-DE 的分辨率和蛋白质的检出率，对 IPG 胶条的改进研究一直在进行，新的胶条不断出现。一方面，胶条的 pH 范围在不断变窄，从最初的 3～10 发展到目前相差 1 个 pH 单位的窄范围 IPG 胶条。这些胶条的出现，不但大大改善了 2-DE 的分辨率，增加了上样量，从而提高了蛋白质的检出，而且可以针对不同样品选择不同的胶条，如酸性蛋白的分离，可选用 pH3～5 或 4～7 的胶条，而碱性蛋白质的分离可选用 pH6～9 的胶条。另外，胶条的 pH 范围也在不断扩大，如 pH3～12 的胶条的出现，改善了碱性蛋白质的分离。

胶条的长度也有不同，有 7 cm、11 cm、13 cm、18 cm 和 24 cm。18 cm 的胶条是目前普遍采用的一向 IEF 胶条，通常可达到 2000 以上蛋白质点的分辨率。这样的分辨能力显然不能满足对细胞中全部蛋白质，包括它们的修饰物进行分离的需要。使用 24 cm 的胶条增加了分离的距离，提高了分辨率。目前已有采用40 cm的胶条进行一向 IEF 分离的研究，结果证明是增加蛋白质分辨率和检出的有效方法[21]。在鼠的蛋白质组研究中，经样品预分离和特定细胞器预分离，采用 40 cm 的 IPG 胶条，共获得 27 752 个蛋白质点，其中一块胶上的蛋白质点数达 10 000 个以上。

4．高通量与自动化

蛋白质组研究的高通量需要研究体系的自动化。如何使 2-DE/MS 为基础的蛋白质组技术体系自动化，是许多研究机构和生物技术公司研究的重点。迄今为止，大量的研究和改进集中在 2-DE 后的样品处理和蛋白质的识别（图 2-3），如用于胶上蛋白质点切割的自动点切割仪，用于胶上蛋白质酶切、回收与点靶的自动样品处理机等。这些仪器的出现，大大减轻了 2-DE 后样品处理的繁复的手工操作，缩短了时间，提高了效率，更重要的是提高了样品处理的重复性，避免了手工操作易产生的污染。另外，一种可同时运行 12 块胶的二向 SDS-PAGE 电泳槽已经研制成功并商品化，如 Ettan Dalt II（Amersham Pharmacia）、Protean II xi (Bio-Rad) 和 Investigator（Genomic Solution），使得二向电泳一次可平行分析的样品数量与一向 IEF 电泳相匹配（一向电泳槽可平行进行 12 个样品的分析），从而大大提高了 2-DE 分析的通量。

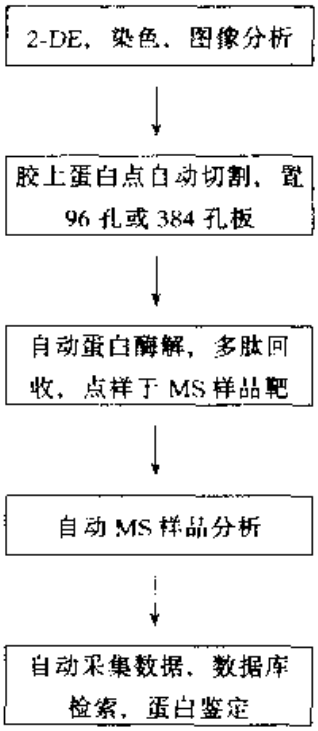


图 2-3　蛋白质识别过程的自动化流程

一种被称为“分子扫描（molecular scanner）”的方法为以 2-DE/MS 为基础的技术体系的自动化打开了思路[22,23]。该方法将一个固相化了胰蛋白酶的 PVDF（polyvinylidene difluoride，PVDF）膜插入 2-DE 电泳胶和另一个 PVDF 膜之间，后一个 PVDF 膜作为收集膜，当对上述体系进行电转印时，胶上的蛋白质在向第一个 PVDF 膜迁移的过程中被胰蛋白酶酶切，酶切肽段穿过带有固相化胰酶的 PVDF 膜，被第二个 PVDF 膜收集。收集膜可直接用于 MALDI-TOF-MS 作肽质量指纹谱分析。该方法实现了 2-DE 电泳后的样品处理与质谱分析一体化，为整个 2-DE/MS 技术体系的自动化打下了基础。该方法的不足之处在于膜上酶切-质谱分析所获得的肽段数目低于目前常规方法所获得的肽段数，使蛋白质特别是翻译后修饰蛋白质的识别与鉴定效率受到影响。

5．2-DE 的局限性

尽管 2-DE 在蛋白质组研究中已获得广泛的应用，但是以 2-DE 为基础的蛋白质组方法体系还存在很多局限。在选择 2-DE 作为主要的蛋白质分离与展示手段时，应当对这些局限有清醒的认识。

（1）实验设计

采用 2-DE 的方法，我们可以获得一个蛋白样品定性和定量的信息。如，通过观察肿瘤病人的细胞蛋白质谱，可以获知哪些蛋白质与肿瘤的发生和发展相关。但是，样本间的异质性以及肿瘤组织中多种细胞的并存，影响实验结果的可

靠性，也是我们在实验设计时必须充分考虑到的。采用激光捕获显微切割的方法（laser-capture microdissection，LCM），可以得到较为单纯的细胞，但是 2-DE 所需的巨大的细胞量（$10^6 \sim 10^7$ 个细胞），是采用 LCM 方法时必须面对的问题。

（2）重复性

2-DE 的重复性尽管得到了很大的改善，但仍然是 2-DE 方法存在的主要问题。同样的操作人员，同样的样品，同样的仪器，几次操作，得到的 2-DE 谱很难一模一样。要想得到好的结果，需要有经验的人员进行操作，并花费大量的时间对 2-DE 的条件进行优化。2-DE 方法的重复性问题，除了电泳条件本身的问题外，样品的获得和预处理的过程也是至关重要的，特别是进行比较蛋白质组研究的时候。

（3）动态范围

除了上述问题外，有限的动态范围与低拷贝蛋白质的难以检测也是 2-DE 目前尚未完全解决的问题。基于 mRNA 的研究方法其动态范围为 5 个数量级，而细胞中蛋白质的动态范围达 7～8 个数量级（细胞中蛋白质的拷贝数可从 $10^1 \sim 10^8$）。采用放射性同位素标记，可以大大提高蛋白质的检测限，但后续的质谱分析的灵敏度难以满足要求。通过加大蛋白质的上样量，可能达到质谱分析的要求，但上样量过大，容易超出电泳胶的负载量，从而降低电泳的分辨率。因此，当采用 2-DE/质谱方法时，要仔细考虑方法的动态范围。

（4）蛋白质的丢失

作为一个真正通用的蛋白质的表达谱展示方法，应当对所有蛋白的分离与检测“一视同仁”，而不应存在歧视效应。2-DE 分析时，部分蛋白质的丢失，特别是疏水性蛋白质和分子质量大于 100 kDa 蛋白质的丢失，使得 2-DE 作为通用的蛋白质分离展示的方法还有一定的局限。尽管对疏水性蛋白质如膜蛋白的提取方法进行了诸多改进，针对酸性或碱性蛋白质的分离也建立了许多方法，有些还相当成功，但是总的来说，2-DE 作为一个通用的分析方法，满足每个实验者的需要还有一定差距。

（5）通量和自动化

尽管对 2-DE/MS 体系的自动化作了很大的改进，特别是 2-DE 胶上蛋白质的识别与鉴定技术基本实现了高通量和自动化，但是，2-DE 分离后的图像分析仍然是整个过程中较费时的步骤，成为整个体系高通量、自动化的瓶颈。

二、色谱-质谱技术

为了弥补以 2-DE/MS 为基础的蛋白质组技术方法的不足，很多实验室致力于开发新的大规模蛋白质分离鉴定技术，并尝试建立新的蛋白质组研究技术平台，如二维色谱-串联质谱技术、亲和色谱-串联质谱等多种模式。

1．二维色谱-串联质谱技术（2D-HPLC-MS-MS）

（1）二维色谱分离蛋白质和多肽

采用液相色谱的方法分离蛋白质和多肽有诸多优点，如速度快，一般几个小时可完成全部分离，而 2-DE 分离一般需 1～2 d；由于在溶液状态样品处理方便、快速，避免了 2-DE 从胶上回收样品的繁复操作，分析过程易于自动化和与质谱联接；对各种蛋白质均适用，包括疏水性、酸性、碱性、分子质量大于 100 kDa、小于 10 kDa 的蛋白质等。2-DE 与 2D-HPLC 方法的比较见表 2-3。

表 2-3　2-DE 与 2-D HPLC 的比较

	优　点	缺　点
2-DE	高分辨率，＞1000 个蛋白质 可得到等电点、相对分子质量信息 可得到蛋白质表达谱 可平行运行多个胶	费时，费力 不易自动化 疏水、酸性、碱性、＜10 kDa、＞200 kDa 蛋白易丢失 样品负载量受限制
2D-HPLC	快速 可分离各种蛋白质 易于自动化 ……样品在液相 ——馏分易收集 可做样品制备或浓缩	不能得到等电点、相对分子质量信息 尚未实现平行操作 不易得到蛋白质表达谱 分辨率不够高

二维色谱分离蛋白质有多种组合模式，第一维色谱一般是离子交换色谱，也有用凝胶过滤色谱，第二维通常是反相色谱。由于反相色谱采用有机溶剂作流动相，避免了盐等添加剂对后续质谱分析的影响。为了提高二维色谱的分离速度和分辨能力，第二维的反相色谱分离系统一般采用较短的色谱柱和较高的流速。整个分离系统采用一根离子交换柱与两根反相色谱柱，通过多通道阀进行柱切换的模式操作（图 2-4）。当蛋白质或多肽组分从离子交换色谱柱上洗脱并进入第一根反相色谱柱时，第二根反相色谱柱上的组分正被洗脱进入检测系统。两根反相色谱柱的轮流切换，使得被分离物可以富集在柱的顶端。适当调整柱切换速度与流速，可以达到每分钟一个循环，一个色谱图。为了减小死体积，采用细内径的管路（0.127 mm）并尽可能减少管路的长度。

采用上述体系分别对标准蛋白混合物和人肺纤维细胞培养上清液蛋白质进行了分离，证明在保留时间、峰高和峰面积保持良好重复性的前提下，峰容量可达到 600[24]。

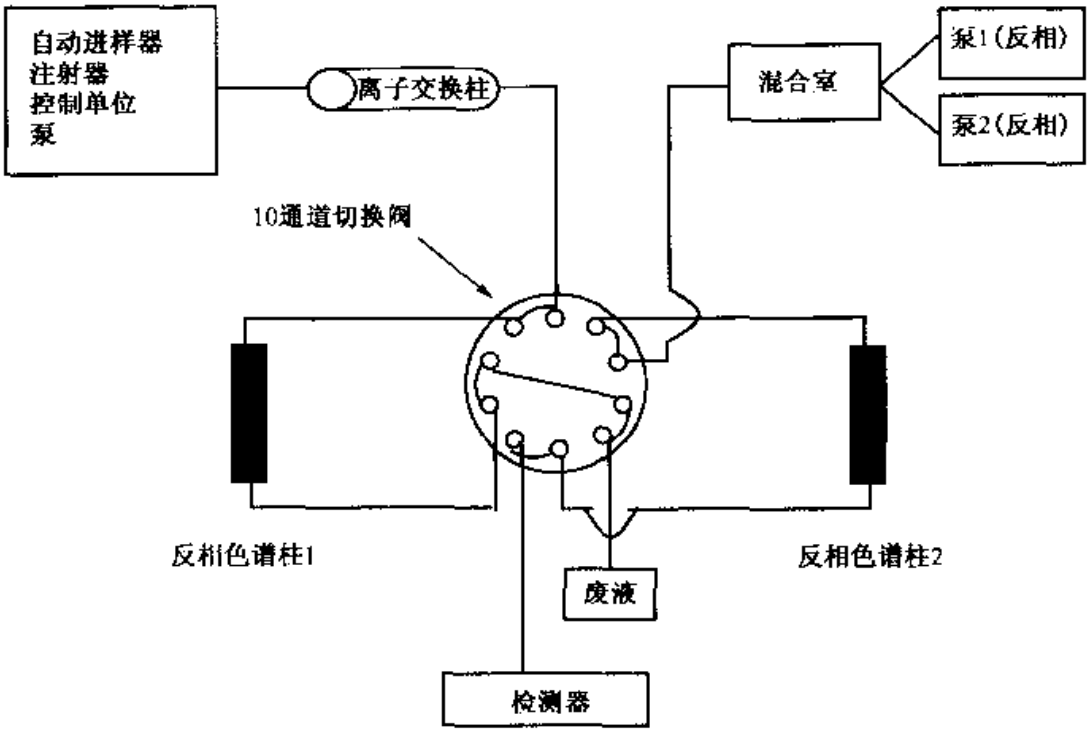


图 2-4　2D-HPLC 分离体系示意图

(2) 二维色谱-串联质谱分离鉴定细胞蛋白质

采用二维色谱-串联质谱技术可以对蛋白质混合物进行直接分离与鉴定。从组织、细胞或其他生物样本获得的蛋白质混合物，用蛋白酶（通常是胰蛋白酶）裂解，得到的多肽混合物样品进行二维色谱分离。质谱分析一般采用电喷雾串联质谱，通过测定肽段的序列对蛋白质进行鉴定。二维色谱-串联质谱（2D-HPLC）分析方法流程见图 2-5。

采用二维色谱/串联质谱（2D-HPLC/MS-MS）大规模分离鉴定蛋白质较成功的例子是酵母蛋白质组的分析[24]。酵母全细胞裂解液经不同强度的缓冲溶液处理，获得可溶、弱溶、难溶三个组分并分别进行蛋白酶解。2D-HPLC 的分离采用内径 100 μm 的毛细管色谱柱，色谱柱先用 5 μm 的反相 C18 填料填充 10 cm，然后用 5 μm 的强阳离子交换填料填充 4 cm。将 420 μg 可溶、440 μg 弱溶和 490 μg 难溶细胞蛋白提取物的多肽混合物分别加样于上述色谱柱中。质谱仪器为 Finnigan 公司的 LCQ 离子阱质谱仪，电喷雾离子源。系统被联接好后，整个分离与鉴定操作自动化进行，采用 4 种不同的缓冲溶液进行 15 步梯度洗脱，质谱数据的分析采用 SEQUEST 算法。

总共 5540 个肽段被指认，1484 个蛋白质被鉴定，其中包括 32 个蛋白激酶，45 个转录因子，131 个跨膜蛋白。这些蛋白质的亚细胞定位分布见表 2-4。在全部 1484 个蛋白质中，有 791 个蛋白质的密码子偏性系数 CBI<0.2，12 个蛋白质

的等电点<4.3，29 个蛋白质的等电点>11，24 个蛋白质的分子质量>190 kDa。与基于 2-DE/MS 的技术平台相比，2D-HPLC/MS 在识别和鉴定低丰度蛋白、膜

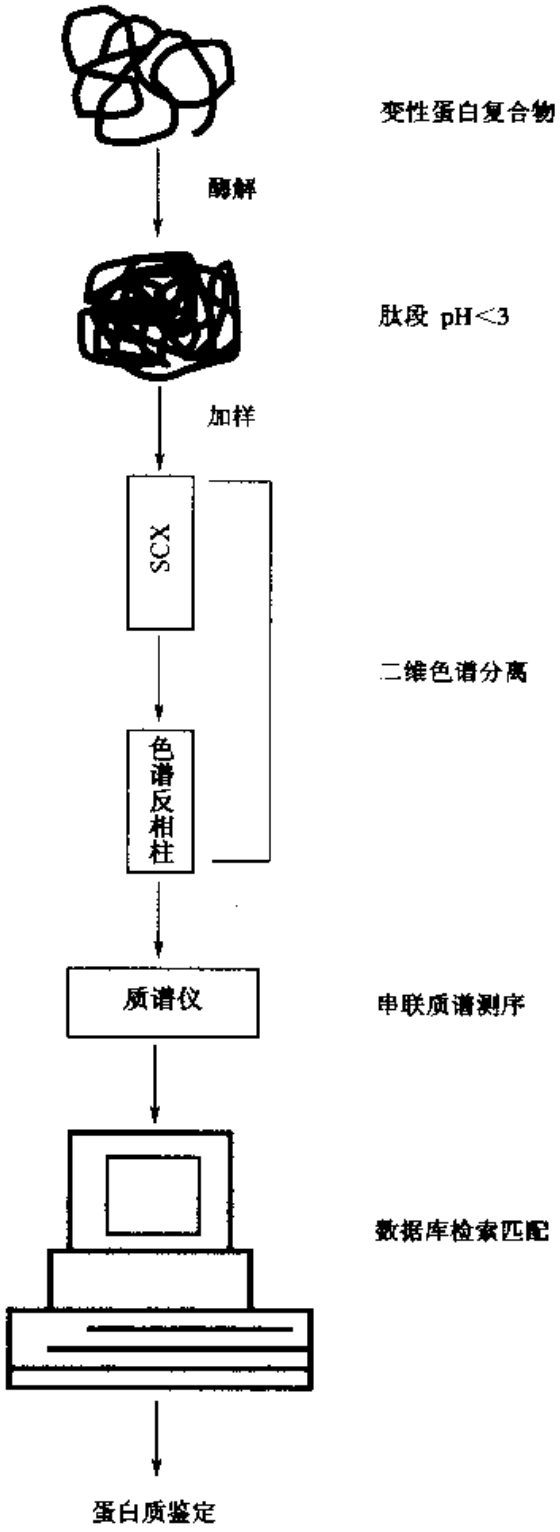


图 2-5　蛋白质混合物二维色谱/质谱分析流程[24]

蛋白、酸性蛋白、碱性蛋白以及相对分子质量大的蛋白质方面有明显的优势。与2-DE/MS技术相比，2D-HPLC/MS方法的不足之处就是通过蛋白质直接酶切得到的肽混合物过于复杂，较难实现对细胞全部蛋白质的识别与鉴定。相信随着对方法的不断改进，2D/HPLC-MS/MS将作为2-DE的补充方法，成为蛋白质组研究中的大规模蛋白质分离与鉴定方法之一。

表2-4　蛋白质的亚细胞定位*

亚细胞器	可溶	弱溶	难溶
细胞壁	2	1	1
质膜	5	18	35
细胞质	286	264	274
细胞骨架	11	20	22
内质网	12	36	42
高尔基体	3	10	16
转运囊泡	4	14	16
核	67	122	151
线粒体	43	87	83
过氧化物酶体	2	3	3
内含体	1	1	2
液泡	5	10	6
微粒体	0	0	1
脂质颗粒	0	2	3

* 啤酒酵母的亚细胞定位依据慕尼黑信息中心蛋白质序列网站；已鉴定的蛋白质中亚细胞定位不清楚的没有列在表中。

2．同位素标记-亲和色谱-质谱技术

为了实现对细胞蛋白质的定量分离与鉴定，一种被称为同位素编码亲和标签的色谱-质谱技术已获得成功应用[25]。人的细胞中约有2万多个蛋白质[25]，当用蛋白酶酶切后，所产生的多肽混合物十分复杂。现有的分离技术，包括2-DE和2D-HPLC都很难对如此复杂的混合物实现完全的分离。采用同位素编码亲和标签技术，不但可以实现对细胞差异表达蛋白质的定量分析，而且大大简化了被分析样品的复杂性，为蛋白质组学技术方法的研究提供了新的思路。

ICAT方法的关键是ICAT试剂的应用。该试剂由三部分组成（图2-6）。中间部分被称为连接部分，分别连接8个氢原子或重氢原子，前者称为轻链试剂，后者称为重链试剂。中间部分一头连接一个巯基的特异反应基团，可与蛋白质中半胱氨酸上的巯基连接，从而实现对蛋白质的标记，另一头连接生物素，用于标

记蛋白质或多肽链的亲和纯化。当采用 ICAT 技术进行蛋白质的定量分离与鉴定时，首先要对蛋白质进行标记，然后进行蛋白酶切，对蛋白质酶切后的多肽混合物进行亲和色谱分离，混合物中仅仅被同位素标记的肽段能够被色谱柱保留，并进入质谱进行鉴定，其他大量肽段都不被色谱柱保留。通过比较标记了重链和轻链试剂肽段在质谱图中信号的强度，可以实现对差异表达蛋白质的定量分析，采用串联质谱技术测定肽段的序列，可以实现蛋白质的鉴定。

同位素编码亲和标签技术利用对标记试剂的巧妙设计，不仅可以实现对不同条件处理的细胞差异蛋白质的定量分析（定量蛋白质组学分析技术详见第八章），而且大大简化了被分离样品的复杂程度，为比较蛋白质组学研究中差异蛋白质的分离与鉴定提供了十分有用的技术。该方法的不足之处在于无法分析和鉴定不含半胱氨酸的蛋白质，另外，目前商品化试剂的高价位也影响了该方法的普遍应用。

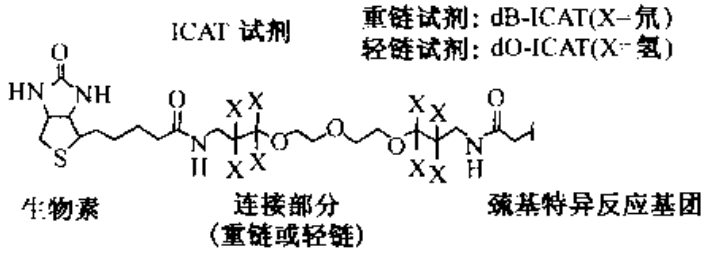


图 2-6　ICAT 试剂[25]

三、一维电泳（色谱）-质谱技术

为了弥补二维电泳的不足，对一维电泳（色谱）-质谱的模式分离鉴定蛋白质复合物或者全细胞蛋白质提取物也进行了尝试。

1．一维 SDS-PAGE 电泳(色谱)-质谱分离鉴定蛋白质复合物

蛋白质复合物的分离与鉴定是功能蛋白质组研究的核心内容。然而，采用传统的 2D-MS 技术有可能造成蛋白质的丢失。SDS-PAGE 电泳结合 MALDI-TOF-MS 技术，或者蛋白酶切结合 LC-MS-MS 技术直接分离和鉴定蛋白质复合物被证明是一种有效的方法。该方法的基本过程见图 2-7。从生物样本中获得的蛋白质混合物，采用免疫沉淀或免疫亲和提取的方法，获得蛋白质复合物。SDS-PAGE 电泳对蛋白质复合物进行分离，染色，胶上原位酶切，MALDI-TOF-MS 分析肽质量指纹谱并鉴定蛋白质，或者对蛋白质复合物直接进行蛋白酶切，对获得的肽混合物进行 LC-MS-MS 分析，通过测定肽片段的序列鉴定蛋白质。

采用上述方法对鼠脑中的 N-甲基化-天冬氨酸受体（N-methyl-D-aspartate receptors，NMDAR）复合物进行了分离鉴定[26]。NMDAR 受体复合物的分离分

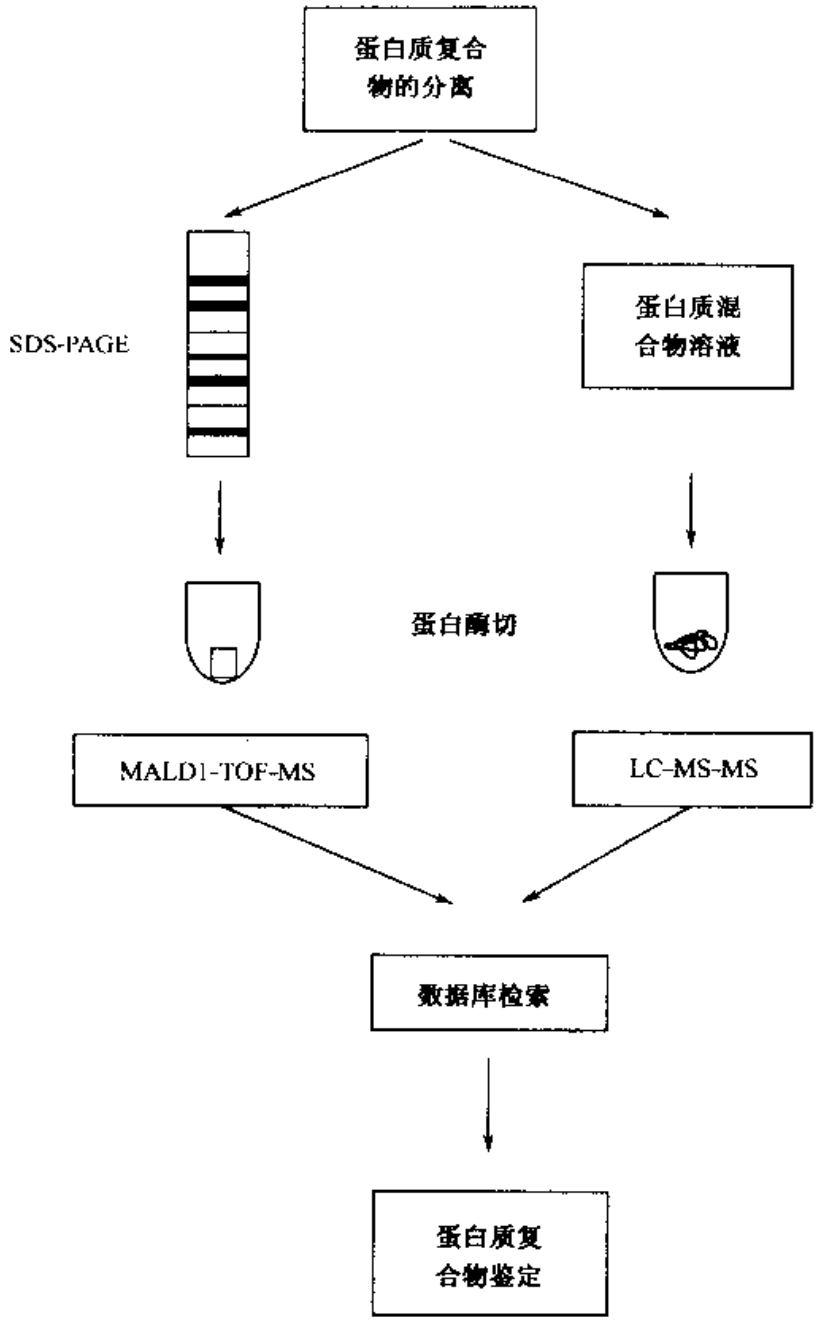


图 2-7　蛋白质复合物一维电泳（色谱）-质谱分析流程图

别采用免疫亲和色谱、抗 NMDAR R1 亚基抗体免疫沉淀和 NMDAR R2B C 端肽与 NMDAR 结合蛋白 PSD-95 亲和结合的方法。实验结果证明，NMDAR 受体复合物由 77 个蛋白质构成，包括神经递质受体、连接物、细胞黏附蛋白、二级信使、细胞骨架蛋白和新蛋白。

2．一维 IEF 电泳-质谱分离鉴定蛋白质

目前采用的以 2D-MS 为基础的蛋白质分析方法，以第一维等电聚焦电泳测定蛋白质的等电点，第二维 SDS-PAGE 电泳测定蛋白质的相对分子质量。与质谱方法相比，SDS-PAGE 测定蛋白质相对分子质量误差大、分辨率差，更重要的是，很难获得翻译后修饰蛋白质相对分子质量变化的准确信息。如何精确测定全细胞蛋白质的相对分子质量，一直是蛋白质组技术方法研究中的难题。一种采用 IPG 胶条进行等电聚焦电泳，采用 MALDI-TOF-MS 对胶条进行表面直接分析的方法已有报道[27]。该方法采用窄范围的 IPG 胶条进行一向等电聚焦电泳，电泳完毕，清洗胶条并用基质溶液浸泡，室温晾干后，将胶条截成 3.5～4 cm 长度，用双面胶固定于质谱的样品靶上并进行 MALDI-TOF-MS 分析。采用 pH5.7～6.0 的 IPG 胶条分离大肠杆菌蛋白质，250 个蛋白质被鉴定，而采用传统的 2D 方法仅鉴定了 100 个蛋白质。测定小于 50 kDa 的蛋白质，相对分子质量准确度为 0.1%～0.2%，pH 准确度在 ±0.3 个单位。该方法免除了样品从 2D 胶向质谱鉴定过渡时的繁复操作，将一向 IEF 电泳与准确、灵敏、高分辨的相对分子质量测定技术相结合，虽然在胶上蛋白酶切和多肽序列的直接测定方面还存在一定问题，但是仍然可能发展成为 2-DE 的替代方法之一，用于蛋白质组的分析。

第三节　高通量蛋白质鉴定技术

对二维凝胶电泳或二维色谱分离的蛋白质，如何进行快速、灵敏、准确的识别与鉴定，是蛋白质组研究所面临的又一重大挑战。传统的蛋白质化学方法，如 Edman 降解或氨基酸组成分析技术，在早期的蛋白质组研究中有所应用[28,29]，但由于 Edman 降解方法的低灵敏度和氨基酸组成分析方法的低准确度，难以满足蛋白质组研究的需要，而已经发展了一百多年的质谱技术，由于其所具有的高灵敏度、高准确度、易于自动化等特点，已经成为蛋白质组研究的关键技术之一。

一、生物质谱技术

自 1886 年 Goldstein 发明早期质谱仪的离子源，1942 年第一台商品化质谱仪的出现，质谱技术的发展走过了一百多年的历程。早期的质谱技术主要应用于地质、环境化学、有机化学和制药等领域，分析对象主要是小分子化合物，质量范围在几千以下。诞生于 20 世纪 80 年代末期的两项软电离质谱技术-电喷雾电离质谱[30]和基质辅助激光解吸电离飞行时间质谱[31]使传统的主要用于小分子化合物研究的质谱技术发生了革命性的变革。ESI-MS 和 MALDI-TOF-MS 所具有的

高灵敏度（amol-pmol）和高质量检测范围（质量数从几万到几十万），使质谱技术真正走入了生命科学的研究领域，并得到迅速发展。

目前商业化的生物质谱仪，其离子化方式主要是电喷雾电离与基质辅助激光解吸电离，前者常采用四极杆质量飞行器，所构成的仪器称为电喷雾（四极杆）质谱仪，后者常采用飞行时间质量分析器，所构成的仪器称为基质辅助激光解吸电离飞行时间质谱仪。电喷雾质谱的特点之一是可以和液相色谱、毛细管电泳等高效分离手段联用，因而可实现复杂化合物分离与鉴定的联机操作。而 MALDI-TOF 质谱仪的特点是对盐和填加物的耐受能力强，且操作简便，测定速度快，易于实现高通量，谱峰与样品成分的质量有一一对应关系，图谱容易解析。此外，用于生物大分子测定的质谱仪还有离子阱（ion trap）质谱仪和傅立叶变换离子回旋共振（fourier transform ion cyclotron resonance，FTICR）质谱等。最近几年出现的新型生物质谱仪是电喷雾-四极杆-飞行时间质谱仪与带有串联质谱功能的 MALDI-TOF 质谱仪。前者在传统的三级四极串联质谱仪的基础上，采用飞行时间质量分析器代替四极杆质量飞行器，大大提高了仪器的分辨率和灵敏度。商品化的仪器有 Q-TOF（英国 Micromass 公司）和 Q-STAR（美国 ABI 公司）等；后者是在质谱中加入源后裂解（post souce decay，PSD）装置或碰撞诱导解离（collisionally induced dissociation，CID）装置，从而使 MALDI-TOF-MS 测定多肽序列成为可能，商品化的仪器有 Reflex III（德国 Bluker 公司）和 Voyager DE STR（美国 ABI 公司）等。

随着蛋白质组研究的深入，又有多种新型质谱仪出现，主要是在上述质谱仪的基础上进行改进与重新组合，本章后面还将详细介绍。

二、2-DE 胶上蛋白质识别与鉴定技术

图 2-8 为 2-DE 胶上蛋白质识别与鉴定流程图。对二维凝胶电泳分离的蛋白质点进行切割，蛋白酶切，提取多肽混合物，采用 MALDI-TOF-MS 首先进行肽质量指纹谱分析[32-36]，约 50%～60%的蛋白质在这一步可以得到鉴定，接着采用电喷雾串联质谱进行肽序列标签测定（peptide sequence tag，PST）[37]，约 20%～30%的蛋白质可以得到鉴定。通过上述两步质谱分析，仍无法得到鉴定的蛋白质，有可能是新蛋白质，需要到核酸数据库中作进一步检索，或者分离、制备一定量，采用全序列测定的方法进行鉴定（有关 MALDI-TOF-MS、ESI-MS-MS 鉴定蛋白质的原理和应用实例详见第六章）。

采用 MALDI-TOF-MS 作为首选方法，对胶上蛋白质进行大规模鉴定的优点是易于实现高通量。蛋白质组研究专用的高通量 MALDI-TOF 质谱仪的出现，以及胶上蛋白质酶解、提取、点靶自动化机器人的商业化，使得 MALDI-TOF-MS 每天分析几百甚至上千个蛋白质成为可能，因而特别适合于胶上蛋白质的高

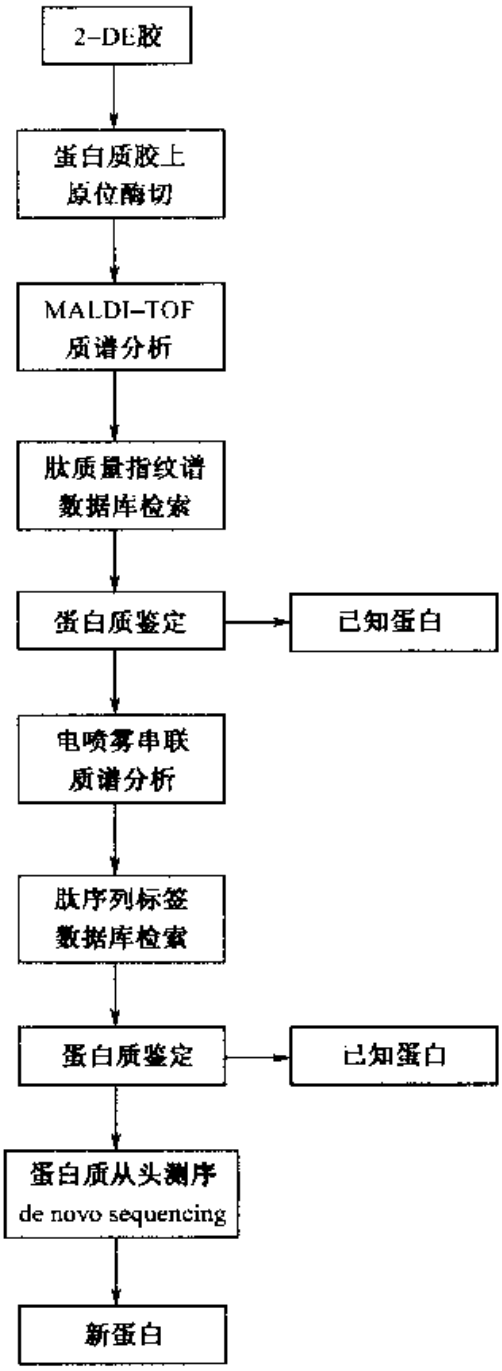


图 2-8　2-DE 胶上蛋白质鉴定流程图

通量、大规模识别与鉴定。但是当所分析的蛋白质样品是混合物，或是数据库中不存在的新蛋白质时，MALDI-TOF-MS 得到的肽质量指纹谱无法与数据库中已有蛋白质的理论肽质量指纹谱进行匹配，因而无法实现对蛋白质的识别与鉴定。另外，由于肽质量指纹谱方法本身的特点，对所使用质谱仪的准确度要求较高，测量误差一般应在 100 ppm 以下，较理想的应是 20～50 ppm，否则分析结果难

以判断，需花费大量时间进行进一步的手工分析，难以实现高通量。这一点在采用高通量 MALDI-TOF-MS 进行自动化分析时显得尤为重要。肽段质量测量准确度与数据库检索结果的相关性见表 2-5。

表 2-5 肽质量准确度与蛋白质数据库检索结果[38]

分子离子	误差（ppm）	匹配蛋白数	
		（所有种类）[a]	（酵母）[b]
1975.954	1000	754	173
	100	200	76
	10	44	6
1562.884	1000	982	120
	100	197	27
	10	19	2
1055.541	1000	1300	173
	100	514	76
	10	87	6

[a]所有种类蛋白质为 52 205 个。[b]酵母蛋白质为 3653 个。

采用电喷雾串联质谱，通过测定肽序列标签对蛋白质进行鉴定的方法，是将质谱测定的多肽序列与数据库中已有蛋白质的序列直接进行比较。与肽质量指纹谱的方法相比，肽序列测定方法不仅特异性高，而且可以实现对混合样品中的蛋白质的识别与鉴定。由于相应的计算机分析软件的发展，对混合样品中混合多肽的序列测定、数据库检索与结果提交完全可以自动化进行。另外，由于电喷雾质谱可以与液相色谱连接，可以实现分离、鉴定一体化操作，因而更加适合复杂样品中蛋白质的识别与鉴定，甚至有可能取代二维凝胶电泳，或作为补充方法，对全细胞蛋白质组分的酶解产物直接进行分离与鉴定[24]。电喷雾串联质谱的不足之处就是分析速度不如 MALDI-TOF-MS 快，另外操作也比 MALDI-TOF-MS 复杂。

三、生物质谱新技术

为了满足高通量、大规模蛋白质组研究的需要，新的生物质谱技术不断出现，如前面提到的电喷雾-四极杆-飞行时间质谱仪以及带串联质谱功能的 MALDI-TOF质谱仪。此外，为了增强 MALDI-TOF 质谱仪的串联质谱功能，从而更有利于实现蛋白质鉴定的高通量，一种带 MALDI 源的四极杆-飞行时间质谱仪和带两个飞行时间质量分析器的 MALDI-TOF-TOF 质谱仪已经出现并商品化。

1．带 MALDI 源的四极杆-飞行时间质谱仪(MALDI QqTOF)

如前所述，由于样品制备简便，操作易自动化，MALDI-TOF-MS 测定肽质量指纹谱的方法已经成为蛋白质组研究中蛋白质鉴定的首选方法。但是，肽质量指纹谱有时不能提供充分的信息以鉴定蛋白质，这时，需要采用 ESI-MS-MS 测定肽序列标签鉴定蛋白质。尽管 ESI-MS-MS 的方法较肽质量指纹谱的方法更加有效，但 ESI-MS-MS 操作费时，不易实现高通量。另外，由于两种质谱分析分开进行，给样品的操作带来麻烦。如果能够实现对同一个样品的 MS 和 MS-MS 分析在同一台设备上进行，无疑将大大提高蛋白质分析鉴定的速度，实现高通量。

MALDI QqTOF 质谱仪的出现，使得上述设想成为了现实。该仪器是将 MALDI 离子源与四极杆-飞行时间-串联质谱仪连接，从而实现了样品靶上的同一个样品在同一台质谱仪上的肽质量指纹谱分析与肽序列标签分析同时进行。目前此种仪器已经商品化，如 MALDI Qstar Pulsar（美国 ABI 公司）和 Q-Tof Ultima GLOBAL（英国 Micromass 公司）。MALDI QqTOF 的结构见图 2-9。

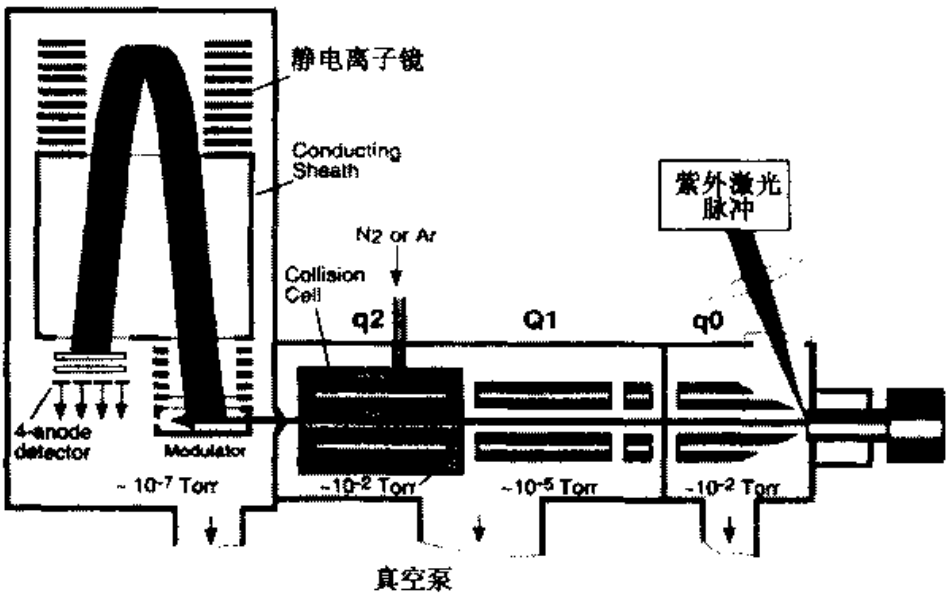


图 2-9　MALDI QqTOF 结构示意图[39]

紫外激光以 20 Hz～50 Hz 的循环速率产生脉冲，脉冲离子束在 q0 碰撞冷却并转换成准-连续流。离子流以不超过 10kHz 的循环速率注射进入质谱仪。

在 MS 操作模式下，由紫外激光脉冲产生的离子流在 q0 碰撞冷却并转换成准-连续流（quasi-continuous beam），接着通过分析四极杆 Q1 到达 TOF 分析器，从而获得全质谱图。在 MS-MS 模式，母离子在 Q1 中被选择，在 q2 中碰撞解离成碎片，产物离子到达 TOF 分析器，从而获得母离子的串联质谱图。肽谱与肽

序列测定在同一块靶上、同一台仪器上进行。由于 MALDI 源产生的离子以直角方式注入 TOF 分析器，解析过程几乎是在完全与质量分析器断开的情况下进行，因而使仪器的校正更加简单和稳定，与传统的轴向模式相比，样品的操作更加方便。更为重要的是，母离子的选择是在注入 TOF 分析器前进行，因而使质谱仪器的性能在单极质谱与串联质谱分析时几乎相同。

MALDI QqTOF 质谱仪已经成功用于 2D 胶上蛋白质的鉴定（图 2-10）[40]。实验结果显示，采用 MALDI QqTOF 质谱仪获得的肽序列信息，经过表达序列标签（expressed sequence tag，EST）数据库检索，成功鉴定出了新蛋白质。

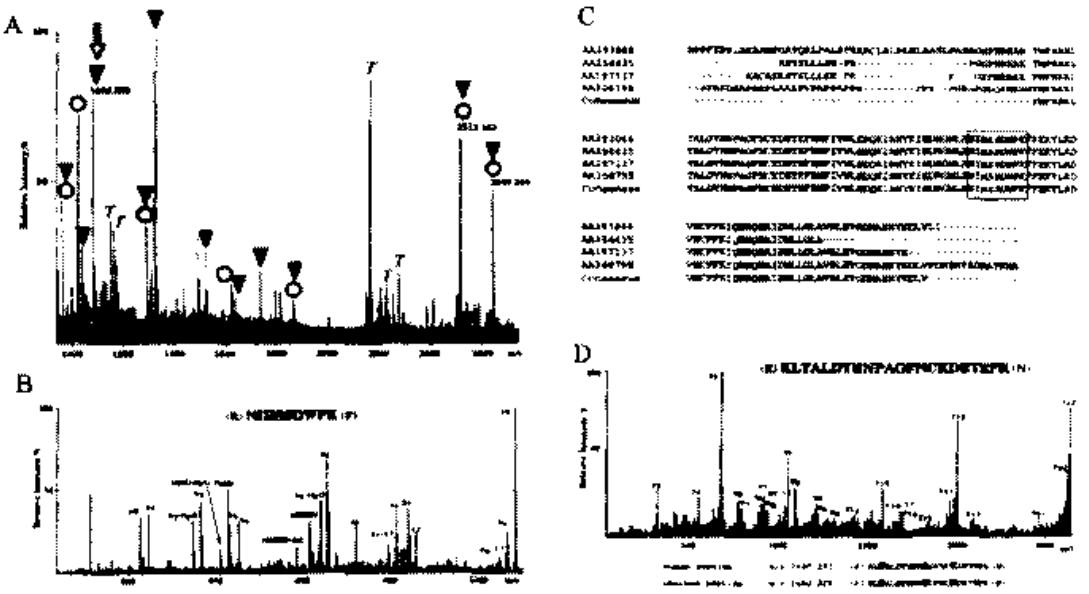


图 2-10　MALDI QqTOF 质谱仪鉴定 2D 胶上蛋白质

图 A 显示胰酶裂解的一个 25 kDa 人蛋白质肽谱。圆圈标记的肽峰与鸟类 CLE7 蛋白匹配。图 B 显示 m/z1083.533（图 A 中箭头指示）的串联质谱图。EST 库检索与（R）NIHSSIDWPK（F）序列匹配。图 C 显示四个人 EST 克隆的同源序列和匹配上的序列（图中方框部分）。10 个肽段离子（图 A 中三角形标记的肽段）与同源序列匹配。图 D 显示母离子 m/z2640.244 的串联质谱图。来源于人和鸟的相应肽段序列列于图下方。

2. MALDI-TOF-TOF 质谱仪

MALDI-TOF-TOF 质谱仪是在传统的 MALDI-TOF 质谱仪的基础上进行改进而成的新型质谱仪。该仪器将 MALDI-TOF 肽质量指纹谱分析的高通量与碰撞诱导解离（collision-induced dissociation，CID）获得的丰富碎片信息相结合，对同一样品先进行肽质量指纹谱分析，接着进行串联质谱分析，使得在微量样品水平上对蛋白质的鉴定在数秒中内完成。与源后裂解测定多肽序列的模式相比（需要在不同质量范围多次采集质谱数据并进行合成），MALDI-TOF-TOF 可以在固定的反射电压下，经过一次采集获得完整的碎片离子质谱图。MALDI-TOF-TOF

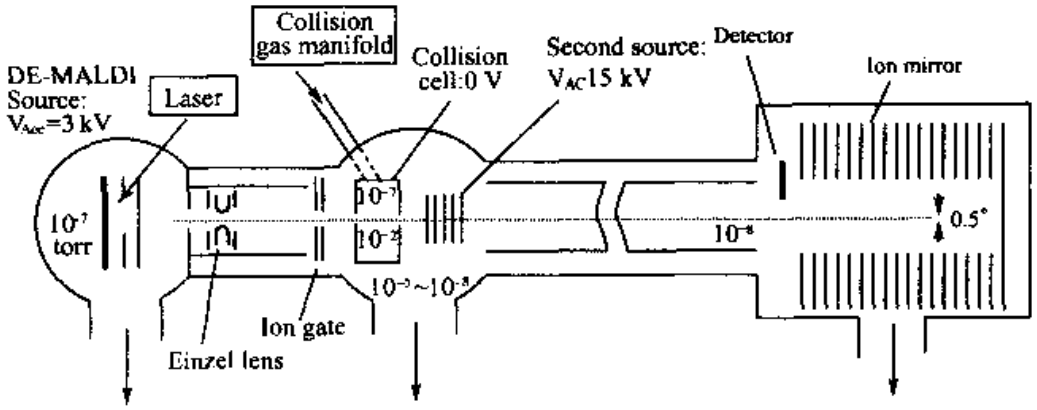


图 2-11　MALDI-TOF-TOF 结构示意图[41]

的结构示意图见图 2-11[41]。目前已经商品化的 MALDI-TOF-TOF 质谱仪有德国 Bruker 公司的 UltraFlex 和美国 ABI 公司的 4700。上述仪器已经成功地应用于电泳胶上银染蛋白质点的鉴定。

四、结语

蛋白质组学的研究开辟了功能基因组学研究的新领域。一方面，蛋白质组研究为探索生命的奥秘提供了崭新的思路和手段，另一方面，蛋白质组研究期待着新的技术方法的突破。随着基因组学与生物信息学技术的发展，政府投资与商业投入的加大，蛋白质组学研究必将在 21 世纪取得突破性的进展。

参考文献

1　Rubin J M. Comparing species. Nature, 2001, 409: 820～821

2　O'Farrell P H. High resolution two-dimensional electrophoresis of proteins. J Biol Chem, 1975, 250: 4007～4021

3　Bjellqvist B, Ek K, Righetti P G, Gianazza E, Gorg A, Westermeier R, Postel W. Isoelectric focusing in immobilized pH gradient: principle, methodology and some applications. J Biochem Biophys Methods, 1982, 6: 317～339

4　Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10 000 daltons. Anal Chem, 1988, 60: 2299～2301

5　Fenn J B, Mann M, Meng C K, Wong S F, Whitehouse C M. Electrospray ionization for mass spectrometry of large biomolecules. Science, 1989, 246: 64～71

6　Ong S E, Pandey A. An evaluation of the use of two-dimensional gel electrophoresis in proteomics. Biomolecular Engineering, 2001, 18: 195～205

7　Gygi S P, Rist B, Gerber S A, Turecek F, Gelb M H, Aebersold R. Quantitative analysis of protein mixture

using isotope coded affinity tags. Nat Biotechnol, 1999, 17: 994－999

8 Currie I, Davison M. Validation and development of fluorescent two-dimensional differential gel electrophoresis proteomics technology. Proteomics, 2001, 1 (3): 377－396

9 Wallach D, Boldin M P, Kovalenko A V, Malinin N L, Mett I L, Camonis J H. Curr Opin Immunol, 1998, 10: 131－136

10 Husi H, Ward M A, Choudhary J S, Blackstock W P, Grant S G N. Proteomic analysis of NMDA receptor-adhesion protein signaling complexes. Nature Neuroscience, 2000, 3 (7): 661－669

11 Link A J, Eng J, Schieltz D M, Carmack E, Mize G J, Morris D R, Garvik B M, Yates J R. Direct analysis of protein complexes using mass spectrometry. Nature Biotechnology, 1999, 17: 676－682

12 Oda Y, Nagasu T, Chait B T. Enrichment analysis of phosphorylated proteins as a tool for probing phosphoproteome. Nat Biotechnol, 2001, 19 (4): 379－382

13 Miriam V, Dwek H A, Anthony R, Leathem J C. Proteome and glycosylation mapping identifies post-translational modification associated with aggressive breast cancer. Proteomics, 2001, 1: 756－762

14 Figeys D, Pinto Devanand P. Proteomics on a chip: promising developments. Electrophoresis, 2001, 22: 208－216

15 Hodges P E, McKee A H Z, Davis B P, Payne W E, Garreis J I. The Yeast Proteome Database (YPD): a model for the organization and presentation of genome-wide functional data. Nucleic Acids Research, 1999, 27 (1): 69－73

16 Klose J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetic, 1975, 26: 231－243

17 Gygi S P, Corthals G L, Zhang Y, Rochon Y, Aebersold R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. PNAS, 2000, 97 (17): 9390－9395

18 兰彦，钱小红，王阁，李勇，罗凌，柳正良，贺福初。蛋白质组分析中蛋白质分步提取方法的建立。生物化学与生物物理进展，2001，28 (3)：415－418

19 Herbert B, Righetti P G. A turning point in proteome analysis: sample prefractionation via multicompartment electrolyser with isoelectric membranes. Electrophoresis, 2000, 21: 3639－3648

20 Klose J. Genotypes and phenotypes. Electrophoresis, 2000, 20: 643－652

21 Bienvenut W V, Sanchez J C, Karmime A, Rouge V, Rose K, Binz P A, Hochstrasser D F. Toward a clinical molecular scanner for proteome research: parallel protein chemical processing before and during western blot. Anal. Chem. 1999, 71: 4800－4807

22 Binz P A, Muller M, Walther D, Bienvenut W V, Gras R, Hoogland C, Bouchet G, Gasteiger E, Fabbretti R, Gay S, Palagi P, Wilkins M R, Rouge V, Tonella L, Paesano S, Rossellat G, Karmime A, Bairoch A, Sanchez J C, Appel R D, Hochstrasser D F, A molecular scanner to automate proteomic research and to display proteome images. Anal Chem, 1999, 71: 4981－4988

23 Wagner K, Racaityte K, Unger K K, Miliotis T, Edholm L E, Bischoff R, Marko-Varga G. Protein mapping by two-dimensional high performance liquid chromatography. J Chromatography A, 2000, 893: 293－305

24 Washburn M P, Wolters D, Yates J R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nature Biotechnology, 2001, 19: 242－247

25 Gygi S P, Rist B, Gerber S A, Turecek F, Gelb M H, Aebersold R. Quantitative analysis of protein mixture using isotope coded affinity tags. Nature Biotechnology, 1999, 17: 994－999

26 Husi H, Ward M A, Choudhary J S, Blackstock W P, Grant S G N. Proteomic analysis of NMDA receptor-adhesion protein signaling complexes. Nature Neuroscience, 2000, 3 (7): 661－669

27 Loo R R O, Cavalcoli J D, VanBogelen R A, Mitchell C, Loo J A, Moldover B, Andrews P C. Virtual 2-D gel electrophoresis: visualization and analysis of the *E. coli* proteome by mass spectrometry. Anal Chem, 2001, 73: 4063～4070

28 Wilkins M R, Gasteiger E, Tonella L. Protein identification with N-and C-terminal sequence tags in proteome projects. J Mol Biol, 1998, 278: 599～608

29 Wilkins M R, Pasquali C, Appel R D, Ou K, Golaz O, Sanchez J C, Yan J X, Gooley A A, Hughes G, Humphery-Smith I, Williams K L, Hochstrasser D F. From proteins to proteomes: Large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Bio/Technology, 1996, 14: 61～65

30 Fenn J B, Mann M, Meng C K, Wong S F, Whitehouse C M. Electrospray ionization for mass spectrometry of large biomolecules. Science, 1989, 246: 64～71

31 Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10 000 daltons. Anal Chem, 1988, 60: 2299～2301

32 Henzel W J, Billeci T M, Stults J T, Wong S C, Grimley C, Watanabe C. Identification proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence database. Proc Natl Acad Sci USA, 1993, 90: 5011～5015

33 James P, Quadroni M, Carafoli E, Gonnet G. Protein identification by mass profile fingerprinting. Biochem Biophys Res Commun, 1993, 195: 58～64

34 Mann M, Hojrup P, Roepstorff P. Use of mass spectrometric molecular weight information to identify proteins in sequence database. Bio Mass Spectrom, 1993, 22: 338～345

35 Pappin D J C, Hojrup P, Bleasby A J. Rapid identification of proteins by peptide-mass fingerprinting. Curr Bio, 1993, 3: 327～332

36 Yates J R III, Speicher S, Griffin P R, Hunkapiller T. Peptide mass maps: a highly information approach to protein identification. Anal Biochem, 1993, 214: 397～408

37 Mann M, Wilm M. Error-tolerant identification of peptides in sequence database by peptide sequence tags. Anal Chem, 1994, 66: 4390～4399

38 Lixin P. A powerful tool for proteomics. China Biomass Shanghai, 2000, p9

39 Loboda A V, Krutchinsky A N, Bromirski M, Ens W, Standing K G. A tandem quadruple/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance. Rapid Communication in Mass Spectrometry, 2000, 14: 1047～1057

40 Shevchenko A, Loboda A V, Shevchenko A, Ens W, Standing K G. MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. Anal Chem, 2000, 72: 2132～2141

41 Medzihradszky K F, Campbell J M, Baldwin M A, Falick A M, Juhasz P, Vestal M L, Burlingame A L. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. Anal Chem, 2000, 72: 552～558

（钱小红）

第三章　二维电泳的蛋白质提取与样品制备

选择合适的样品制备方法对获得满意的二维电泳图谱是很重要的。由于蛋白样品的来源和类型差异较大，因此，在这里仅提供常用的裂解策略。对每一种样品而言，合适的裂解步骤由经验而定，目的是尽可能地完全溶解、解离、变性和还原蛋白。

当进行样品制备时，很重要的一点是明确其研究目的。是获得尽可能多的蛋白，还是仅获得所感兴趣的某些蛋白？全蛋白表达谱或可重复的清晰图谱哪一个更重要？尽管采用一些附加的样品制备方法会提高二维电泳图谱的质量，但是同时也会导致某些种类蛋白的丢失。因此，必须谨慎地权衡上述几者之间的关系来做出决定。

对不同类型的蛋白进行样品制备时，需要采用不同的方法和条件。某些蛋白在天然状态下与细胞膜、核酸或其他蛋白形成复合物；某些蛋白形成各种非特异性聚合体；而某些蛋白在脱离其正常环境时则发生沉淀。溶解的效果依靠选择细胞破碎方法、蛋白解聚和溶解方法、去污剂和裂解液成分等方法来实现。如果其中任何一个步骤没有得到优化，分离很可能是不完全的，或者是变异的。

为了完全分析所有的细胞内蛋白，细胞必须进行有效的破碎。破碎方法的选择依赖样品是否来源于细胞、固体组织或者其他的生物材料，同时也依赖于这种分析是获得所有的蛋白，还是特殊的亚细胞器组分。一些温和的和剧烈的裂解方法将在下面进行讨论。

蛋白酶在细胞破碎时很可能释放出来，其导致的蛋白降解将使二维电泳图谱的分析复杂化，因此蛋白样品应该利用蛋白酶抑制剂加以保护。

如果组织中只有一类蛋白质是有意义的，那么在样品制备过程中进行分级分离是必需的。如果从某一特殊的亚细胞器（细胞核、线粒体和质膜）中提取蛋白，那么在进行二维电泳之前，首先应该通过分级分离或其他的方法进行纯化。

沉淀样品中的蛋白，清除杂质是可选择的步骤，这主要依赖于样品本身和研究目的。它可以除去样品中的盐离子、小分子、离子去污剂、核酸、多糖、脂类和酚类等杂质。通常建议采用最简单的样品制备方法。例如，具有低蛋白浓度和高盐的样品可以被正常地稀释、分析或脱盐，随后通过冰冻干燥进行聚集，或者通过三氯醋酸、冰丙酮沉淀，再以重泡胀溶液（rehydration solution）溶解。通常利用重泡胀溶液稀释样品已经足够，如果蛋白聚集和杂质难以去除，则沉淀和

清除步骤是必需的。

样品制备原则：

1. 尽可能采用简单方法进行样品处理，以避免蛋白丢失。
2. 细胞和组织样品的制备应尽可能减少蛋白的降解，低温和蛋白酶抑制剂可以防止蛋白的降解。
3. 样品裂解液应该新鲜制备，并且分装冻存于 -80 ℃。勿反复冻融已制备好的样品。
4. 通过超速离心清除所有的杂质。
5. 加入尿素之后加温不要超过 37 ℃，防止氨甲酰化而修饰蛋白。

第一节　细胞裂解的方法

一、温和的裂解方法

通常应用于组成比较简单的样品，例如组织培养的细胞、血细胞和一些微生物，或者用于分析某一特定的细胞器，例如只需要裂解胞质蛋白、完整的线粒体或其他的细胞器。有时这些技术联合起来应用（在蛋白酶抑制剂存在下进行渗透裂解、在去污剂存在下进行反复冻融）。

1) 渗透裂解[1]：主要应用于血细胞、组织培养细胞；通常用低渗溶液悬浮细胞。
2) 冻融裂解[2-4]：常用于细菌、组织培养细胞；通常用液氮迅速冷冻细胞，随后在 37 ℃融化，反复几次。
3) 去污剂裂解：用去污剂溶解细胞膜、裂解细胞，释放其内容物；常用于组织细胞；通常用含有去污剂的裂解液重悬细胞，也可以直接用样品裂解液或重泡胀溶液进行裂解。另外，如果阴离子去污剂 SDS 用于裂解细胞，可以用含有兼性离子去污剂、非离子去污剂的裂解液稀释样品，也可以用丙酮沉淀以分离出 SDS。
4) 酶裂解法[5,6]：某些细胞含有细胞壁，需要首先用酶来消化细胞壁。例如：用溶菌酶（lysozyme）来消化细菌的细胞壁；用纤维素酶（cellulase）和果胶酶（pectinase）消化植物；用溶细胞酶（lyticase）消化酵母细胞壁。此法常用于植物、细菌和真菌。通常这些酶在等渗溶液中处理细胞。

二、剧烈的蛋白裂解

常用于难于破碎的细胞，如固体组织内的细胞，或具有坚硬细胞壁的细胞，如酵母细胞。

1) 超声裂解法[7~9]：超声仪可产生超声波，通过切应力来裂解细胞。在剧烈搅拌的状态下会产生剪切效果，但需注意应最大可能减少热量和气泡的产生。此法常用于细胞样品。通常用迅速的超声重悬细胞，并在冰上冷却。

2) 弗氏压碎器（French Pressure cell)[5,6,10]：在高压下迫使细胞穿过小孔径而产生剪切力，从而裂解细胞。此法常用于含有细胞壁的微生物，如细菌、酵母和藻类。通常将细胞悬液置于预冷弗氏压碎器中，加压后收集粗提液。

3) 研磨法[7,11,12,13]：用研钵或研杵研磨细胞，此法常用于固体组织、微生物。组织和细胞通常冻存于液氮中，随后研磨成粉状。铝或砂有助于研磨。

4) 机械匀浆法[1,2,14~16]：常用于固体组织。Dounce 和 Potter-Elvehjem 匀浆器可用于破碎细胞或一些软组织；而混合器或一些研磨设备用于一些较大的组织。通常将组织切成小片，加入预冷的匀浆缓冲液（3～5 倍于组织体积），简单匀浆，通过过滤或离心获得裂解液。

5) 玻璃珠匀浆法[5,6,17]：通常剧烈振荡的玻璃珠可以打破细胞壁，释放细胞内容物。此法常用于细胞悬液或微生物。用等体积的预冷裂解液重悬细胞，置于结实的管子里。每克细胞（湿重）加入 1～3 g 玻璃珠，剧烈振荡 1 min，冰浴 1 min。重复上述步骤。

三、用蛋白酶抑制剂防止蛋白裂解

当进行细胞裂解时，蛋白酶会释放出来。这会严重影响二维电泳的结果，因此需要采用一些策略。如果可能的话，直接加入强变性剂（8mol/L urea，10% TCA 或 2% SDS)。蛋白酶在低温下无活性，因此尽可能在低温下进行样品制备。另外，许多组织蛋白酶在 pH 超过 9.0 的时候失活，因此在样品裂解液中出现 Tris、碳酸钠和载体两性电解质时也可以抑制蛋白降解。虽然这些方法可防止蛋白降解，但是有些蛋白酶在上述条件下仍然保持降解蛋白的活性。此时，需要应用蛋白酶抑制剂。每一种蛋白酶抑制剂仅对特殊的一类蛋白酶发生作用，因此建议使用“蛋白酶抑制剂鸡尾酒（protease inhibitor cocktail)”。

1) PMSF（苯甲基磺酰氟)：是一种不可逆抑制剂，常用浓度为 1 mmol/L，灭活丝氨酸蛋白酶和一些半胱氨酸蛋白酶，但是其局限在于可在水溶液中迅速失活，因此，现用现加效果较好；另外，在二硫苏糖醇（DTT）或者 β-巯基乙醇溶液中 PMSF 的抑制效果可能会减少，因此，含巯基的试剂可在晚些时候加入。

2) AEBSF（Pefabloc^{TM} SC)：通常工作浓度为 4 mmol/L，作用与 PMSF 相

似，但溶解性较好，且毒性低。但是 AEBSF 可能改变蛋白的等电点。

3) 金属蛋白酶抑制剂乙二胺四乙酸、乙二醇双四乙酸（EDTA、EGTA）：通常应用 1mmol/L 的工作浓度，通过螯合自由的金属离子来抑制金属蛋白酶活性。

4) 肽蛋白酶抑制剂：亮抑酶肽（leupeptin），它是可逆性蛋白酶抑制剂，在含 DDT 的溶液中以低浓度的形式保持活性，抑制一些丝氨酸和半胱氨酸蛋白酶的活性；胃蛋白酶抑制剂 A（pepstatin A），抑制天冬氨酸蛋白酶（如酸性蛋白酶）的活性；抑蛋白酶肽（aprotinin），抑制许多丝氨酸蛋白酶；苯丁抑制素（bestatin），抑制氨基蛋白酶。但是，这些蛋白酶抑制剂价格昂贵，而且它们是小肽片段，可能会出现在二维电泳图谱中，其中胃蛋白酶抑制剂 A 在 pH＝9 的时候不能抑制蛋白酶。

5) 甲苯磺酰赖氨酰氯甲酮（TLCK），甲苯磺酰苯丙氨酰氯甲酮（TPCK）：通常作用浓度在 0.1～0.15mmol/L，这些相同的成分不可逆抑制许多丝氨酸蛋白酶和半胱氨酸蛋白酶。

6) Benzamidine：通常作用浓度为 1～2mmol/L，抑制丝氨酸蛋白酶。

四、蛋白沉淀步骤

这是一种可选择的方法。通常用于选择性清除杂质，如盐离子、核酸、脂类等会干扰二维电泳结果。重悬之后沉淀蛋白也可以用来制备浓度稀释的样品（如植物、尿液）。沉淀并不是完全有效的，某些蛋白在沉淀后并不容易重悬。因此，在样品制备的过程中，应用沉淀的步骤可能会改变蛋白的二维电泳图谱。

1．硫酸铵沉淀（盐析）

原理：在高盐溶液中，蛋白质倾向于聚合，并从溶液中沉淀下来。许多潜在的杂质（如核酸）将保持在溶液中。

步骤：蛋白浓度大于 1 mg/ml，缓冲液浓度大于 50 mmol/L，并含有EDTA，缓慢加入硫酸铵至饱和，搅拌 10～30 min，通过离心沉淀蛋白[18]。

然而许多蛋白在高盐溶液中是可溶的，因此，这种方法只能被用来预分离或富集蛋白。并且，残存的硫酸铵会干扰等电聚焦（IEF），必须被清除。

2．TCA 沉淀（三氯醋酸沉淀）

这是一种有效的蛋白沉淀方法，将 TCA 加到提取液中，终浓度高达 10%～20%。蛋白质可于冰上沉淀 30 min。另外，组织样品可以直接用 10%～20%的 TCA 进行匀浆。这种方法可限制蛋白降解和其他的化学修饰。最后用丙酮或乙醇清洗沉淀，以除去 TCA[19,20]。

然而，蛋白质再溶解是很困难的，并且不能完全再溶。残留的 TCA 必须通过丙酮或乙醇进行彻底清洗，若过多暴露于低 pH 溶液中可能导致某些蛋白降解或修饰。

3．丙酮沉淀

这种有机溶剂通常用来沉淀蛋白，以清除许多杂质，如去污剂、脂类。

于提取物中加入至少 3 倍体积的冰丙酮，使蛋白在 -20 ℃沉淀至少 2 h，再离心沉淀蛋白。残留的丙酮通过空气干燥或冻干除去[21~24]。

4．在丙酮中用 TCA 沉淀

两者联合应用更加有效，通常用 10% TCA 在丙酮中重悬裂解的样品溶液（含有 0.01%β-巯基乙醇溶液或 20mmol/L DTT）。至少在 -20 ℃沉淀 45 min，通过离心沉淀蛋白，并用含有 0.01%β-巯基乙醇溶液或 20mmol/L DTT 的冰丙酮清洗沉淀。通过空气干燥或冻干去除残留的丙酮[7,12,25~28]。

此法仍然存在难再溶的现象，同 2。

5．用苯酚提取，随后在甲醇中用醋酸铵沉淀

这对于杂质含量高的植物样品很有效。

蛋白质被提取到饱和酚中，然后在甲醇中用醋酸铵从酚相中沉淀蛋白。这些沉淀在甲醇中用醋酸铵洗涤几次，然后用丙酮清洗，残留的丙酮通过蒸发除去[20~26,29,30]。

但此法较复杂，并且耗时较多。

五、清除影响二维电泳图谱的杂质

样品中非蛋白的不纯物质可能干扰蛋白的分离及二维电泳图谱的质量，因此，在样品制备中需考虑清除这些杂质。

1．盐离子、痕量缓冲液和其他带电荷的小分子

盐离子可干扰电泳过程，应尽可能保持低浓度或加以清除。在 IPG 胶条中，盐离子可导致溶液的电导率增大。通过延长 IEF 的聚焦时间，直到离子移至胶条末端时聚焦才会完成。水的运动也可能导致胶条一端干燥，一端膨胀，在胶条中的盐离子可能导致较大区域不能聚焦（如水平条纹）。

如果 IPG 胶条在样品中进行重泡胀，则重泡胀溶液的盐浓度应低于 10 mmol/L。如果在样品杯中加样，则可耐受 50 mmol/L 左右的盐浓度。然而，当蛋白忽然移到一个低盐环境中时，可能在加样处发生沉淀。

通常用透析、旋转透析、凝胶过滤和沉淀/重悬的方法进行脱盐处理。

透析是一个很有效的方法，可出现较少的样品丢失，然而耗时较多，并要求较大体积的溶液。尽管旋转透析较迅速，但蛋白吸附于透析膜也是一个问题，它应该在加入尿素和去污剂之前应用。凝胶过滤和沉淀/ 重悬后同样可导致蛋白丢失。

2．内源性小分子（如核酸、代谢物和磷脂）

这些物质通常带负电荷，能发生偏阳极侧的聚焦不全。

通常 TCA/丙酮沉淀在除去这类物质时特别有效。

3．离子去污剂

离子去污剂，如 SDS，通常用于蛋白的提取和溶解，但却强烈干扰 IEF。SDS 与多肽形成复合物。如果 SDS 不被除去，则不能正确聚焦。

通常用含有兼性离子去污剂或非离子去污剂的重泡胀溶液稀释含 SDS 的样品，因此，其终浓度为 0.25%或更低。其他去污剂与之的比率为 8:1。

通过丙酮沉淀也可去除部分 SDS。在高温下沉淀将最大限度除去 SDS。但是在－20 ℃沉淀会更完全[21]。

4．核酸（RNA、DNA）

核酸增加样品的黏度，并导致背景弥散；高分子质量的核酸能阻滞凝胶孔径；核酸可能通过电稳态相互作用与蛋白结合，从而阻碍聚焦；如果分离的蛋白通过银染检测，凝胶中的核酸也将染色，导致高背景。

通常可用 DNase/RNase A 混合物处理样品中的核酸，将其变为单或寡核苷酸。这经常需加入 0.1 倍体积的含有 1mg/ml DNase、0.25mg/ml RNase A 的溶液和 50mmol/L $MgCl_2$ 溶液在冰上孵育[14,24]。

然而 DNase 和 RNase A 也可能出现在二维电泳图像中。

超离也可用来除去大分子核酸，然而，也会除去高相对分子质量蛋白。当应用低离子强度的提取溶液时，带负电荷的核酸很可能与带正电荷的蛋白形成复合物。高离子强度提取或高 pH 提取可最大程度减少这些相互作用（注意：提取时所加的盐离子必须在后续步骤中除去）。

5．多糖

多糖能阻碍凝胶孔径，或导致沉淀，或延长聚焦时间，或出现水平条纹。某些多糖含有负电荷，能与蛋白质形成复合物。

通常用硫酸铵或苯酚/ 醋酸铵沉淀的方法，随后离心。

超速离心可以除去高分子多糖。

6. 脂类

许多蛋白质，尤其是膜蛋白，可以与脂类形成复合物，这会降低其溶解性，影响其等电点和分子量。脂类也能与去污剂形成复合物，减低其效率。

通常用强变性剂和去污剂最大程度减少蛋白与脂类的相互作用，但可能需过多的去污剂。

7. 酚类组分

苯酚通常在许多植物组织中出现，通过酶催化的氧化反应来修饰蛋白[23,26]。

通常在组织提取过程中应用还原剂防止苯酚的氧化（如β-巯基乙醇、DTT等），也可通过沉淀方法迅速从酚组分中分离蛋白。或者用抑制剂，如硫脲来灭活多酚氧化酶。或用 PVP 或 PVPP 吸收酚来清除酚组分。

8. 不溶物质

在样品中的不溶物质能阻碍凝胶孔径，并导致聚焦不良。当应用样品杯时，它能阻碍蛋白进入 IPG 胶条中。

通常在 IEF 之前，应除去不溶物质。

六、裂解液的组分

为完成聚焦良好的第一向等电聚焦，蛋白必须完全溶解和解聚。无论样品是相对的粗提物还是经过沉淀步骤，样品裂解液必须包含一定的组分以确保在进行第一向等电聚焦之前的完全溶解和变性。基本成分包括：尿素、一种或几种去污剂。目前，含有尿素（urea）和兼性离子去污剂如 CHAPS 的裂解液已经被发现具有良好的裂解效果。还原剂和 IPG 缓冲液（IPG buffer）也可以增强样品的溶解性（如图 3-1 所示）。

在变性条件下进行等电聚焦可以给予最高的分辨率和最清晰的图像。尿素作为一种中性离液剂，通常在裂解液中的浓度为 8 mol/L。尿素可溶解和解折叠大多数的蛋白质，形成完全随机的构象，使所有的离子基团暴露于溶液中。近来，硫脲的加入可以进一步提高溶解度，特别是膜蛋白。

在裂解液中均包含非离子去污剂和兼性离子去污剂，以确保蛋白完全溶解和防止通过疏水相互作用导致蛋白聚合。最初，两个相似的非离子去污剂 NP-40 和 Triton X-100 被广泛应用。随后研究表明：兼性离子去污剂 CHAPS 的效果更好，基本浓度达 4%。如果上述溶解发生困难，还可以考虑用 SDS，但由于它是一种阴性离子去污剂，所以须与其他的去污剂合用，并且其终浓度应低于 0.25%，或者其他去污剂与 SDS 的比率为 8:1。

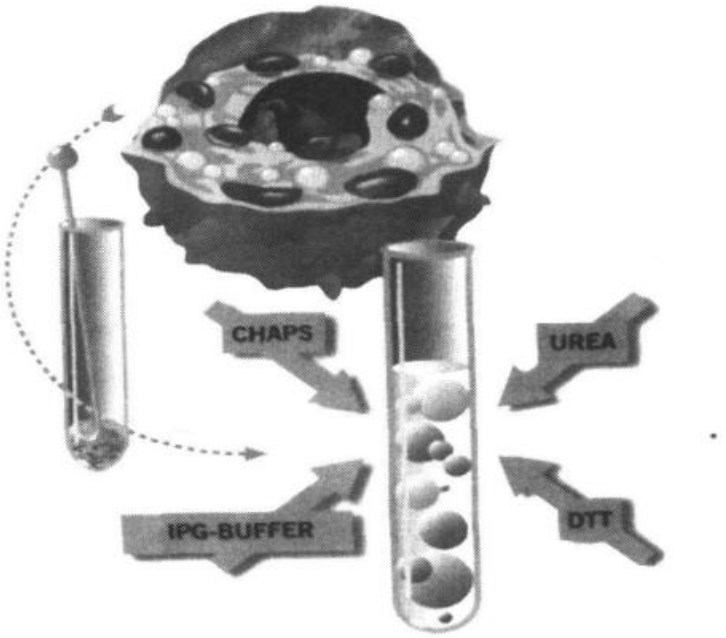


图 3-1　细胞裂解液的组分示意图

还原剂可以断裂二硫键，以保持蛋白处于一种还原状态。起初，β-巯基乙醇被用来作为一种还原剂，但它不仅要求高浓度，而且某些不纯的物质容易产生人工假像。目前，最常用的还原剂是二硫苏糖醇（DTT），浓度变化范围在 20～100mmol/L 之间。二硫赤藓糖醇（DTE）与 DTT 相似，也可以作为还原剂应用。近来，非巯基的还原剂三丁基膦（tributyl phosphine，TBP，2mmol/L）也被用来作为还原剂。

载体两性电解质（carrier ampholytes）和 IPG buffer 也包含于裂解液中，可以通过电荷与电荷之间的相互作用减少蛋白聚合来增强其溶解性。

样品在进行离心之前应该在室温下保持至少 1 h 来达到完全的变性和溶解。在含有尿素的溶液中不可加热，防止蛋白的修饰作用。超声裂解可以帮助加速溶解。

现根据样品的类型不同简要进行介绍。

1．可溶性样品

可溶性液体样品如血清、血浆、尿样、脑脊液以及细胞和组织的水溶性提取物，经很少的前处理便可进行二维电泳分析。含蛋白质浓度较高的样品，如血浆和血清，可用适量样品裂解缓冲液稀释后直接分析。但是，诸如白蛋白和免疫球蛋白一类的高丰度蛋白质在 2-DE 图谱上常干扰其他表达量少的成分。将这些高丰度蛋白质消除可以克服这个问题，但同时可能导致非特异性蛋白质丢失[31,32]。

其他类型的液态样品可能含有较低的蛋白质浓度或较高盐浓度，这将会干扰

IEF 时蛋白质的分离，这种样品可在二维电泳分离前用透析或液相色谱脱盐，样品可以通过冻干、聚乙二醇的透析、超滤或用 TCA/丙酮沉淀的方法进行浓缩。研究表明，TCA/丙酮沉淀可导致蛋白酶失活，进而减少蛋白质降解，同时可以去除杂质，富集极碱性蛋白质，如从全细胞裂解液中富集核糖体蛋白[28]。

2．组织样品

固体组织样品常在裂解缓冲液中破碎，最好的方法是将组织在液氮中冷冻后破碎。包裹在铝箔纸并贮存于液氮中的较小组织样品，可以用研杵和研钵在液氮环境下压碎，大的组织块可在溶解缓冲液中用旋转叶片型匀浆器进行匀浆处理，但尽量要避免加热和泡沫的产生。

组织样品存在的一个特殊问题是样品异质性。组织样品和来源于疾病病灶的样品包含许多种不同类型的细胞，分别收集组织的病灶区和非病灶区难度较大。几种用于二维电泳分析的方法可解决这个问题，通常用基于免疫亲和技术的正性或负性选择，以富集特殊的细胞群，这些样品经裂解后可直接应用于 2-DE 分析，或是先进行短期的原代细胞培养之后再通过二维电泳分析。目前，微量切割技术可用来从组织切片中获取纯的细胞群，其中最受关注的是激光捕获微量切割技术（LCM）[33]。首先，在倒置显微镜下将激光束聚焦在所选择的组织切片区域，用激光束活化与组织切片接触的转移膜来选择细胞。激光束准确击中的区域转移膜与紧贴着的组织结合在一起，因此可以将这些组织中的细胞从周围组织中提取出来。若用 PCR 技术扩增这些微量标本中的 DNA 或 RNA，则使 LCM 技术很容易地应用到核酸分析中。与此对应的，缺乏相应的蛋白质放大手段将是蛋白质组分析中 LCM 分离细胞所遇到的主要挑战。而 Banks 等则证实了 LCM 在蛋白质组分析中的应用，在他们的研究中，通过对一例正常人宫颈表皮组织切割分析，与整个组织进行分析对比，结果显示一些蛋白质可由 LCM 而得到富集[34]。

3．细胞

对体外悬浮培养生长的细胞或周期性细胞样品（如红细胞、淋巴细胞），理想的方法是通过离心收集，随后用磷酸盐缓冲液清洗并溶于样品缓冲液。对于在固体基质中培养的细胞，应首先去除培养基质，用 PBS 或等渗蔗糖溶液清洗细胞层，后者可特别有效地减少可能干扰一向等电聚焦的盐离子，然后通过刮擦方法收获细胞，但应避免使用蛋白裂解酶，或者，加入少量体积裂解缓冲液，将附着在基质上的细胞直接裂解。如果细胞核酸含量过高而导致样品黏度太大时，可用不含蛋白质酶的 DNase/RNase 混合物来处理样品[35]。

4．样品分级

由于真核细胞蛋白质表达的动态范围高和多样性，有时必须对蛋白质进行分

级处理，以降低样品复杂性并富集低拷贝蛋白质。蛋白质预分级可以用亚细胞收集、液相电泳、吸附色谱和选择性沉淀等方法来实现[36]。此外，还有一种方法是根据蛋白质在一系列溶解能力逐渐增强的缓冲液中分步溶解性能不同而实现对细胞或组织蛋白质的分级提取[37,38]。

第二节　蛋白质的分步提取技术

在目前蛋白质组领域中，所鉴定的蛋白质组通常仅占预期蛋白质组的一小部分。虽然对细胞器和质膜预分级可减少样品的复杂性，然而，这要求相当多的经验和价格昂贵的仪器，如超速离心机。有人提出利用蛋白溶解性的不同分步提取蛋白。此法现已被用于提取各种不同的组织，其中通常采用以水相、有机相和表面活性剂为基础的提取溶液。但是，仍然很难提取到样品的全部蛋白质组和低丰度蛋白。此外，一些疏水蛋白，包括膜蛋白和膜相关蛋白，以及高度抗性组织如头发和皮肤组织的蛋白几乎很难进行二维电泳分离，特别是在应用固相 pH 梯度的 IEF 中。近来，有许多文献报道一些试剂在等电聚焦之前可以大大提高样品的溶解度，如硫脲、sulfobetaine（一种表面活性剂）和 TBP。通过联合硫脲、表面活性剂 SB3-10 和不带电荷还原剂 TBP，可产生一个高效的 IEF 样品裂解溶液；而且，还可以溶解那些传统方法中不溶的蛋白。然而，这种增强溶解性的缺陷在于，对于复杂样品而言，二维电泳的分离会出现许多重叠的蛋白斑点，特别是对上样量较大的制备型电泳。这使得 Molloy 等学者在 1998 年将分步提取法和高效裂解缓冲液联合起来，拓展到 IEF 中，以获得高分辨率的、清晰的二维电泳图谱。具体操作步骤如下：

第一步——水溶液的提取：约 5～15 mg 细胞加入 2ml 裂解液 I（40mmol/L Tris）。反复冻融 3～4 个循环，加入适量的 DNase I 和 RNase A，振荡混匀。离心收集上清液，冷冻干燥器中抽干后即为第一步提取物。

第二步——含 Urea/CHAPS/DTT 溶液的提取：向第一步的沉淀中加入 500 μl的裂解液 II（8 mol/L Urea，4% CHAPS，100 mmol/L DTT，40 mmol/L Tris，0.5% Pharmalyte 3-10，20 μg/ml DNase I 和 5 μg/ml RNaseA）。剧烈振荡混匀，离心收集上清即为第二步提取物。

第三步——含硫脲/SB3-10/TBP 溶液的提取：将上一步所余的沉淀用 40 mmol/L Tris 清洗后，加入 200 μl 裂解液 III（5 mol/L Urea，2 mol/L thiourea，2% CHAPS，2% SB3-10，2mmol/L TBP，40mmol/L Tris，0.5% Pharmalyte 3-10，20 μg/ml DNase I 和 5 μg/ml RNase A）。强烈振荡混匀，离心后收集上清。保存于－70 ℃（见图 3-2[39]）。

分步提取法所采用裂解液的溶解性能是逐渐增加的。第一步裂解液只含有 40mmol/L Tris，其余为去离子水，只溶解偏亲水性的蛋白质。第二步裂解液属

于传统的裂解方法，含有表面活性剂 CHAPS、还原剂 DTT、解聚剂 Urea 等成分，由于具有良好的溶解能力，可以溶解亲水性、中性和较疏水性的蛋白。第三步裂解液中添加了能够促进疏水蛋白质溶解的成分，如表面活性剂 SB3-10、还

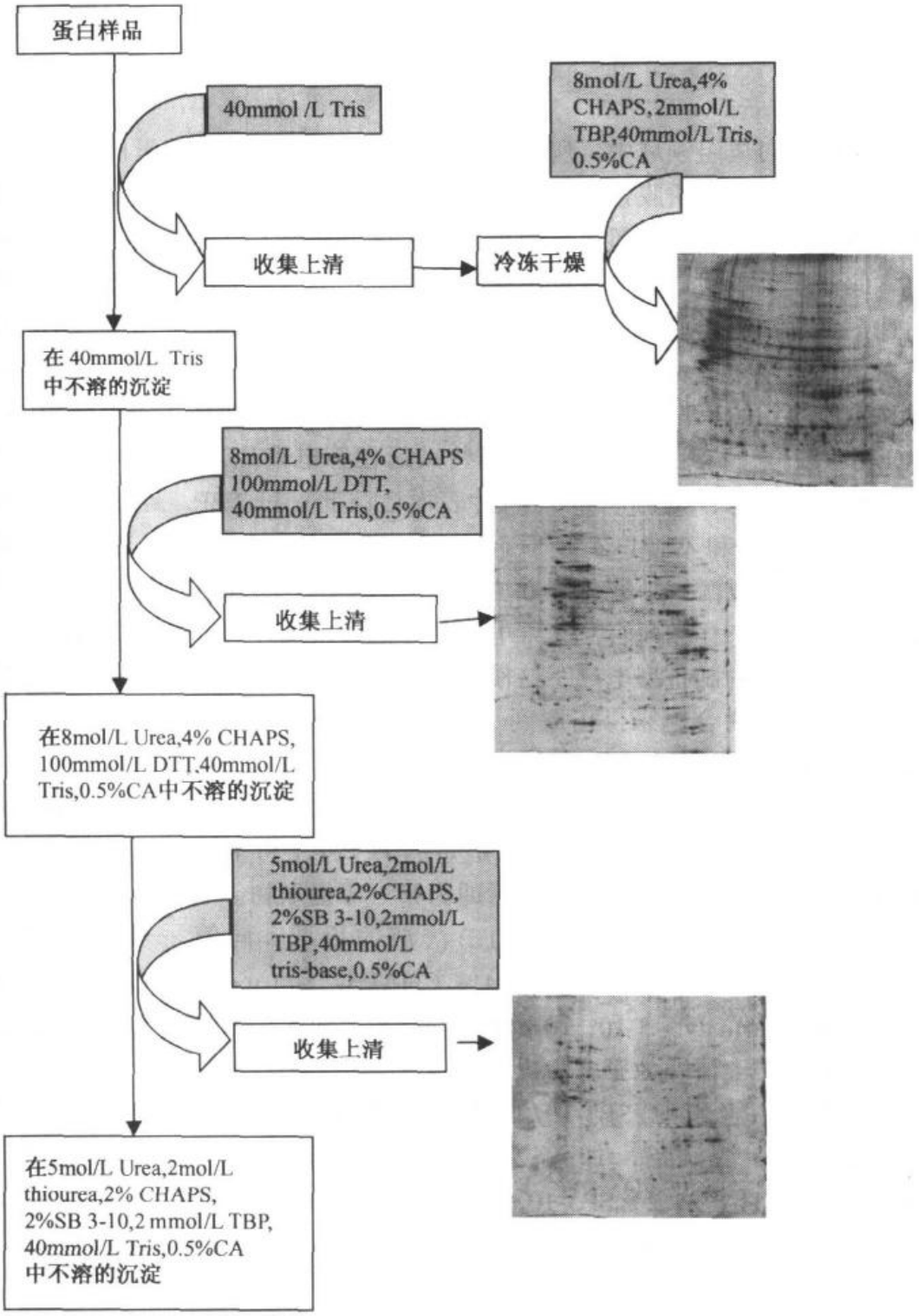


图 3-2　分步提取法示意图

原剂 DTT、解聚剂 thiourea，主要溶解偏疏水性的蛋白质[40]。有报道表明，11个大肠杆菌膜蛋白从这最后一步的裂解液中鉴定出来[39]。

上述基本的操作步骤可依据样品类型而变化。在多数情况下，大部分蛋白质在第一步和第二步操作中可被提取，而且第一步和第二步操作很可能出现相似的图谱。在提取过程中的重叠现象依赖于破碎细胞和组织所选用的物理过程的效率高低。例如，在超声破碎时，大肠杆菌在数秒内即可裂解完全，而结核杆菌则需25 min 才达到 90%的裂解。这些信息需要在进行分步提取之前就了解清楚，以便获得最优化的裂解效果。

这种分步提取方法已证实是一种快速有效的分级和浓缩不溶性蛋白（如膜蛋白）的方法。进而，分步提取蛋白与细胞器分级联合起来也被证明是高效的方法。例如：分步提取纯化的细胞器不但可以提供细胞的定位，而且还可以通过分离可溶性组分导致二维电泳图谱的简化。在蛋白质组研究领域中其发展目标之一是：将继续优化预分级和增强蛋白的溶解度，并联合多范围 pH 梯度进行二维电泳的分离，以期获得更高的分辨率，获得更多的蛋白质组，提供更丰富的生物学数据。

第三节　亚细胞分离与蛋白质提取

亚细胞器的纯化和分级可获得高度纯化和富集的蛋白。最基本的方法是依据蛋白的大小和密度差异来进行亚细胞分级。这些方法的第一步是在低速离心力的作用下，从胞质细胞器中分离出细胞核和未破碎的细胞，获得所需上清。随后，通过各种密度梯度从上清中分离各种细胞器，如线粒体、溶酶体等。Huber 等人在 1996 年研究原癌基因转化对内体细胞器的影响时就采用了上述的方法。研究表明：经亚细胞分级，膜分级组分在内体富集，随后进行二维电泳分离，计算机辅助图像分析显示出差异的蛋白。另外，前述已提及由于新的去污剂、表面活性剂和离液剂的出现，使溶解效果显著增加，并联合分步提取蛋白的方法，通过二维电泳技术已经成功分离出 12 个跨膜蛋白[41]。若联系某些分离蛋白的特殊定位，则对帮助了解蛋白的功能具有一定的提示作用[42]。此外，细胞、组织可被处理分离成其他特殊的组分，随后用二维电泳加以分析。这些细胞分级包括：细胞核[43]、核糖体[44,45]、细菌细胞壁[46,47]、叶绿体[48]和胞外基质上清[49]。

这部分内容主要以膜蛋白、核蛋白为例，简要介绍亚细胞的蛋白质提取。其余部分请参见后面章节。

一、膜蛋白的提取和分离

膜蛋白被简单定义为与膜，例如脂双分子层，相关的蛋白。然而，“相关”

二字似乎意味着不确定性。从结构角度而言，膜蛋白指多肽链跨越脂双分子层数次的蛋白，这些蛋白被定义为膜整合或膜内在蛋白。由于脂双分子层的低电环境限制了跨膜氨基酸的结构排组，所以，膜内在蛋白的跨膜结构域必须折叠形成α螺旋或β片层的二级结构。近来，从几个基因组的全序列估计，跨膜蛋白（锚定到膜上的蛋白除外）占细胞总蛋白的30%[50,51]。另外，膜蛋白处于细胞的边界，在细胞的各种生命活动中发挥重要的作用，如信号转导、细胞黏附和代谢、离子交换、内吞等。因此，它们也可以作为细胞药物治疗的靶点。

在进行膜蛋白提取的时候，有两点需要注意。首先，膜蛋白的纯度如何？当细胞在水相、去污剂中进行裂解时，细胞膜和内膜网络将会断裂成碎片或颗粒，通过沉淀或分级技术加以分离。同时，这些技术也能分离膜连的细胞器（如线粒体）。然而，在这种样品制备中，并不是每一个蛋白都是膜蛋白。到目前为止，实际上多数关于膜蛋白的研究在进行样品制备时或多或少地含有膜样品制备的粗提物。并没有任何证据表明，在二维电泳图谱所呈现的斑点中是真正的膜蛋白，而不是可溶性蛋白杂质。文献报道，通过生化方法已经产生了两种广泛应用的技术可最低限度地减少样品制备的污染。一种是应用盐，主要是溴化钾[52]，或更多的离液剂进行清洗，可以使与膜相互作用较弱的蛋白分离出来（亦可尝试应用碳酸钠），目的是富集膜内在蛋白。从而，在二维电泳的凝胶图谱中尽量使膜整合蛋白占主要部分。另一种方法是应用去污剂分级来除去膜样品制备中的可溶性杂质[53]。膜蛋白通常位于两相去污剂系统中相对较丰富的那一相中，但是，不是所有分配于此相的蛋白都是膜蛋白。某些去污剂的种类、温度和离子强度将影响膜蛋白的分级效率。

另外，无论是动物细胞、组织、酵母、细菌和植物，膜蛋白很难出现在二维电泳凝胶图谱中，原因是：膜蛋白是低丰度的、多为偏碱性、难溶于等电聚焦的水相介质中的蛋白。因此，为了能够从二维电泳凝胶图谱中分离出膜蛋白，首先必须实现3个条件：① 必须保证膜的脂类环境，同时，应避免多余的脂类对等电聚焦造成干扰；② 必须以一种可溶的形式（通常是去污剂-蛋白复合物）从膜中分离出来；③ 所提取的膜蛋白必须在等电聚焦过程中维持可溶状态，特别是在等电点的位置。尽管这些条件限制了用二维电泳分析膜蛋白，但是，根据现在技术的发展，这种限制是可以解决的。例如：现在的制备型电泳可以使上样量达到毫克级；而且，新的、宽范围的pH预制干胶条的出现对此也提供了一定的帮助，使之变得容易解决。并且，文献报道一些新的去污剂、离液剂、还原剂及有机溶剂能够溶解膜内在蛋白。例如：Chevallet M报道利用相分离技术从拟南芥叶中纯化质膜蛋白[54]。Rabilloud T等利用4个不同的步骤（Triton X-100、Triton X-114、碳酸盐处理和氯仿/甲醇处理）来检验从质膜中提取疏水蛋白的能力，利用6个不同的裂解缓冲液来检验溶解疏水蛋白的能力。裂解缓冲液包括：7 mol/L Urea，2 mol/L thiourea，0.5% Triton X-100，1.2% Pharmalytes 3-10，20

mmol/L DTT 和 2% W/V 新去污剂或 4% CHAPS。新去污剂包括 ϕC5、ϕC6、ϕC7、ϕC8、ASB14。这项研究表明，能溶解疏水蛋白的新去污剂的使用依赖于样品中的脂含量多少[55]。依据上述方法已经通过质谱鉴定出一系列相关的膜内在蛋白（见表 3-1[56]）。然而，与疏水相关的问题仍很难解决。很明显，研究目的是建立一种蛋白疏水性与二维电泳两者相容的上限。尽管新的去污剂、离液剂、还原剂及有机溶剂的应用，即前述所提到的分步提取蛋白的方法可以分离出富集的膜蛋白[57]，并且，硫酸盐类的去污剂似乎比含有 PEG 或糖基头的去污剂更有效，但常规线性的硫酸盐类与二维电泳所必需的离液剂的相容性很弱[58]。虽可应用各种化学组分（如氨基）以增加溶解性，但这些组分又可能降低去污剂的作用[59]。总之，生物膜蛋白的二维电泳需要依据样品的不同要求预先优化溶解条件，才能达到更完美的分离。

表 3-1　在二维电泳凝胶图谱上已鉴定的来源于动物和组织细胞的膜内在蛋白

蛋白	TMD	相对分子质量 /kDa	GRAVY	TMD/kDa	2-D gel technology[d]
CYP1A1	1	58	−0.084	0.017	CA/U/NID
fgrl	1	89.5	−0.8	0.011	IPG/TU/C
Fc receptor	1	40	−0.34	0.025	CA/U/NID
CD2[a]	1	39	−0.89	0.025	CNU/NID
CD25[a]	1	28	−0.68	0.035	CA/U/NID
CD45[a]	1	144	−0.63	0.007	CA/U/NID
Integrin α5	1	110	−0.221	0.009	CA/U/NID
Integrin β1	1	86	−0.458	0.011	CA/U/NID
Stomatin	1	31.6	0.04	0.03	IPG/TU/ZW
Transferrin receptor	1	85	−0.24	0.01	CA/U/NID
HLA-D	1	25	0.029	0.04	CNU/NID
NMDA receptor[a]	3	103	−0.3	0.03	CA/U/TU/C
GABA α receptor[b]	4	48	0.76	0.08	CNU/NID
Angiotensin rec.[b]	7	41	0.51	0.17	CNU/NID
ACTH receptor[a]	7	33	0.76	0.205	CA/U/NID
Bradykinin receptor[b]	7	44	0.46	0.16	CNU/NID
Parathormone rec[a]	7	62	0.11	0.113	CA/U/NID
Band Ⅲ	12	101.8	0.2	0.11	IPG/TU/ZW
Glucose transporter[b,c]	12	54	0.53	0.22	IPG/U/C

[a]仅通过抗体鉴定。

[b]通过蛋白纯化分析。

[c]仅通过等电聚焦。

[d]CA, carrier ampholytes IEF; IPG, immobilized pH gradient; U, Urea; TU, Urea + thiourea; NID, non-ionic detergent (NP-40 or Triton X-100); C, CHAPS; ZW, zwitterionic detergents。

二、核蛋白的提取和分离

由核膜包裹的细胞核含有遗传信息，它提供基因表达的空间。在成纤维细胞中，细胞核占据细胞容量的5%。因为许多细胞核蛋白属于中等溶解度的，所以联合尿素、硫脲和去污剂裂解（2mol/L 硫脲、5mol/L 尿素和2% SB3-10）核蛋白，可达到足够的分辨率。尽管细胞核蛋白是强碱性蛋白（pI>10），通常与二维电泳的第一维等电聚焦相容程度较低[28]，但在涉及凋亡的研究中，已经建立好从 Burkitt 淋巴瘤中纯化的核蛋白的二维电泳图谱[60]。利用蛋白质组学的方法建立的细胞核蛋白质组，多参考 SWISS-2DPAGE 内的人肝脏细胞核[61]，可以从 ExPASy 服务器上获得(http://www.expasy.ch/ch2d)。而在 SWISS-PROT 蛋白数据库中，已含有 964 个人细胞核蛋白[62]。另外，与细胞核蛋白相关的某些数据库在1993年已被提及[63]。目前，英国爱丁堡的 MRC 人类遗传学研究中心准备启动建立一个关于细胞核蛋白的特殊数据库，可以通过国际互连网络进行搜索查询(http://www.hug.mrc.ac.uk/users/Wendy.Bickmore/npdintro.html)。

核基质是 Berczney 和 Coffey 于 1974 年提出的概念，为非染色质结构蛋白及细胞核经高盐离子、核酸酶、去污剂抽提后所剩的核蛋白网状结构。核基质包括核表面核纤层蛋白、多种低丰度核蛋白、核内不均一核糖核蛋白及与基质相关的 DNA 结合区。通常采用 Fey 和 Penman 建立的核基质蛋白提取方法。即：在含有 0.5% Triton X-100 和 1.2mmol/L 蛋白酶抑制剂的缓冲液中匀浆数分钟，离心去除脂类和可溶性蛋白。其沉淀与提取液（100mmol/L KCl，3mmol/L $MgCl_2$，0.5% Triton X-100，1.2mmol/L PMSF）于 4 ℃反应 15 min，离心去除可溶性骨架蛋白。沉淀再于消化液内（50mmol/L NaCl，3mmol/L $MgCl_2$，1.2 mmol/L PMSF，100 μg/μl DNase，100 μg/μl RNase A）室温消化 30 min，离心后沉淀蛋白主要包括核基质蛋白和中间丝。沉淀蛋白溶于 8 mol/L 尿素的蛋白裂解液内后，经透析除去尿素。不溶性细胞外基质、糖及中间丝，经超速离心沉淀下来，上清即为细胞核基质蛋白[64]。Gerner C 等人应用纳喷-ESI-MS/MS 已经鉴定出33个蛋白，主要属于核基质蛋白[65]。

真核细胞核膜组成包括：与内质网连续的核外膜、由核纤层支撑的核内膜、由核纤层蛋白组成的中间纤维网络和位于核孔复合体的连接内外核膜的核孔膜。核外膜的蛋白成分与内质网相似。迄今为止，几乎没有特殊的蛋白与内膜相关；核纤层蛋白 B 受体及其相关的多肽、Emerin、Nurim 和 MAN-1 等蛋白或是位于核纤层上，或是与染色质相结合，参与染色体的结构，这对于维持细胞核的正常形态与功能是必需的[66]。核膜含有几个富集在不同组分的膜整合蛋白和多蛋白复合物，而且许多膜整合蛋白与结构蛋白的相互作用网络不能通过去污剂提取，这个问题与任何其他复合物的膜蛋白的分析相似。通常采用不同的提取策略来完

成核膜蛋白的分级制备，经过 MALDI-TOF 已经鉴定出 129 个已知蛋白。其中，已鉴定的 19 个基因产物此前在蛋白水平并没有被检测到。具体的核蛋白及核膜蛋白的提取和分级制备参照 Dreger M 的方法。简而言之，细胞核用冰 TP 缓冲液［10mmol/L Tris-HCl（pH7.4），10mmol/L NaH_2PO_4（pH7.4），1 mmol/L PMSF，10 μg/ml aprotinin，10 μg/ml leupeptin］匀浆，缓冲液中含有 250 μg/ml heparin，1 mmol/L Na_3VO_4，10 mmol/L NaF，400 U Benzon nuclease。在 4 ℃下搅拌 90 min，细胞核膜通过离心（10 000×g，30 min，4 ℃）沉淀下来，随后用 STM 0.25 缓冲液［20 mmol/L Tris-HCl（pH7.4），0.25 mol/L Sucrose，5 mmol/L $MgSO_4$，1 mmol/L Na_3VO_4，1 mmol/L PMSF，10 μg/ml aprotinin，10 μg/ml leupeptin］重悬沉淀。蛋白浓度用 Bradford 方法检测。为了进一步分级，在 STM 0.25 缓冲液（每个样品约有 200 μg 蛋白）中的核膜提取物或者加入 Triton X-100（TX-100；终浓度为 0.5%）、或者加入 Urea 和 Na_2CO_3（终浓度各为 4 mol/L 和 0.1 mol/L）、或者加入 NaCl（终浓度为 1 mol/L）。样品在 4 ℃下振荡 15 min。TX-100 的抗性物质以及经 NaCl 清洗的样品在 13 000 r/min 的转速下离心获得。离液剂抗性物质在 50 000 r/min 转速下超速离心可以获得。蛋白浓度同样用 Bradford 方法检测。这种方法一方面可以减少样品的复杂性，因此可以弥补聚丙烯酰胺凝胶的分辨率的问题，另一方面，不同核膜蛋白的分级可以含有来自于既定核膜蛋白成分特征的一组已知蛋白，这些蛋白提供了研究未知蛋白或非特征性蛋白的结构网络的基础，并可预测其功能[67]。

综上所述，随着人类基因组测序的结束，蛋白质组的研究将会到达一个新的高度，这将激励人们寻找更好的方法进行样品制备和蛋白提取（如分离提取膜蛋白和低丰度蛋白），为深入进行蛋白质组研究奠定基础[57]。

参考文献

1 Dignam J D. Preparation of extracts from higher eukaryotes. Methods Enzymol, 1990, 182: 194－203

2 Lenstra J A, Bloemendal H. Topography of the total protein population from cultured cells upon fractionation by chemical extractions. Eur J Biochem, 1983, 135: 413－423

3 Bollag D M, Edelstein S J. Protein Extraction. In: Protein Methods. NY: Wiley-Liss 1991

4 Toda T, Ishijima Y, Matsushita H, Yoshida M, Kimura N. Detection of thymopoietin-responsive proteins in nude mouse spleen cells by two-dimensional polyacrylamide gel electrophoresis and image processing. Electrophoresis, 1994, 15: 984－987

5 Cull M, McHenry C S. Preparation of extracts from prokaryotes. Methods Enzymol, 1990, 182: 147－153

6 Jazwinski S M. Preparation of extracts from yeast. Methods Enzymol, 1990, 182: 154－174

7 Görg A, Postel W, Günther S. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis, 1988, 9: 531－546

8 Kawaguchi S I, Kuramitsu S. Separation of heat-stable proteins from *Thermus thermophilus* HB8 by two-dimensional electrophoresis. Electrophoresis, 1995, 16: 1060－1066

9 Teixeira-Gomes A P, Cloeckaert A, Bezard G, Dubray G, Zygmunt M S. Mapping and identification of Brucella melitensis proteins by two-dimensional electrophoresis and microsequencing. Electrophoresis, 1997, 18: 156～162

10 Ames G F L, Nikaido K. Two-dimensional gel electrophoresis of membrane proteins. Biochemistry, 1976, 15: 616～623

11 Görg A, Boguth G, Obermaier C, Posch A, Weiss W. Two-dimensional polyacrylamide gel electrophoresis with immobilized pH gradients in the first dimension (IPG-DALT): The state of the art and the controversy of vertical versus horizontal systems. Electrophoresis, 1995, 16: 1079～1086

12 Görg A, Postel W, Domscheit A, Günther S. Two-dimensional electrophoresis with immobilized pH gradients of leaf proteins from barley (*Hordeum vulgare*): Method, reproducibility, and genetic aspects. Electrophoresis, 1988, 9: 681～692

13 Posch A, van den Berg B M, Burg H C J, Görg A. Genetic variability of carrot seed proteins analyzed by one- and two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis, 1995, 16: 1312～1316

14 Geigenheimer P. Preparation of extracts from plants. Methods Enzymol, 1990, 182: 174～193

15 Theillet C, Delpeyroux F, Fiszman M, Reigner P, Esnault R. Influence of the excision shock on the protein metabolism of *Vicia faba* L. meristematic root cells. Planta, 1982, 155: 478～485

16 Wolpert T J, Dunkle L D. Alternations in gene expression in sorghum induced by the host-specific toxin from *Periconia circinata*. Proc Natl Acad Sci USA, 1983, 80: 6576～6580

17 Blomberg A, Blomberg L, Norbeck J, Fey S J, Larsen P M, Larsen M, Roepstorff P, Degand H, Boutry M, Posch A, Görg A. Interlaboratory reproducibility of yeast protein patterns analyzed by immobilized pH gradient two-dimensional gel electrophoresis. Electrophoresis, 1995, 16: 1935～1945

18 Englard S, Seifter S. Precipitation techniques. Methods Enzymol, 1990, 182: 285～300

19 Wu F S, Wang M Y. Extraction of proteins for sodium dodecyl sulfate-polyacrylamide gel electrophoresis from protease-rich plant tissues. Anal Biochem, 1984, 139: 100～103

20 Meyer Y, Grosset J, Chartier Y, Cleyet-Marel J C. Preparation by two-dimensional electrophoresis of proteins for antibody production: antibodies against proteins whose synthesis is reduced by auxin in tobacco mesophyll protoplasts. Electrophoresis, 1988, 9: 704～712

21 Guy G R, Philip R, Tan Y H. Analysis of cellular phosphoproteins by two-dimensional gel electrophoresis: applications for cell signaling in normal and cancer cells. Electrophoresis, 1994, 15: 417～440

22 Halloway P, Arundel P. High-resolution two-dimensional electrophoreisis of plant proteins. Anal Biochem, 1988, 172, 8～15

23 Flengsrud R, Kobro G. A method for two-dimensional electrophoreisis of proteins from green plant tissues. Anal Biochem, 1989, 177: 33～36

24 Matsui N M, Smith D M, Clauser K R, Fichmann J, Andrews L E, Sullivan C M, Burlingame A L, Epstein L B. Immobilized pH gradient two-dimensional gel electrophoresis and mass spectrometric identification of cytokine-regulated proteins in ME-180 cervical carcinoma cells. Electrophoresis, 1997, 18: 409～417

25 Damerval C, de Vienne D, Zivy M, Thiellement H. Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. Electrophoresis, 1986, 7: 52～54

26 Granier, F. Extraction of plant proteins for two-dimensional electrophoresis. Electrophoresis, 1988, 9: 712～718

27 Tsugita A, Kamo M, Kawakami T, Ohki Y. Two-dimensional electrophoresis of plant proteins and standard-

ization of gel patterns. Electrophoresis, 1996, 17: 855～865

28 Görg A, Obermaier C, Boguth G, Csordas A, Díaz J J, Madjar J J. Very alkaline immobilized pH gradients for two-dimensional electrophoresis of ribosomal and nuclear proteins. Electrophoresis, 1997, 18: 328～337

29 Hurkman W J, Tanaka C K Solubilization of plant membrane proteins for analysis by two-dimensional gel electrophoresis. Plant Physiol, 1986, 81: 802～806

30 Usuda H, Shimogawara K. Phosphate deficiency in maize. VI. Changes in the two-dimensional electrophoretic patterns of soluble proteins from second leaf blades associated with induced senescence. Plant Cell Physiol, 1995, 36: 1149～1155

31 Dunn M J. Two-dimensional polyacrylamide gel electrophoresis. In: Advances in Electrophoresis (ed. Chrambach. A., Dunn, M. J. and Radola, B. J.). Vol. 1. Weinheim. VCH, 1987, 1～109

32 Lollo B A, Harvey S, Liao J, Stevens A C, Wagenknecht R, Sayen R S, Vhale J, Sajjadi F G. Improved two-dimensional gel electrophoresis representation of serum proteins by using ProtoClear. Electrophoresis, 1999, 20: 854～859

33 Simone N L, Bonner R F, Gillespie J W, Emmert-Buck M R, Liotta L A. Laser-capture microdissection: opening the microscopic frontier to molecular analysis. Trends Genetic, 1998, 14: 272～276

34 Banks R E, Dunn M J, Forbes M A, Stanley A, Pappin D, Naven T, Gough M, Harnden P, Selby, P J. The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis-preliminary findings. Electrophoresis, 1999, 20: 689～700

35 Garrels J I. Two dimensional gel electrophoresis and computer analysis of proteins synthesized by clonal cell lines. J Biol Chem, 1979, 254: 7961～7977

36 Corthals G L, Wasinge V C, Hochstrasser D F, Sanchez J C. The dynamic range of protein expression: A challenge for proteomic research. Electrophoresis, 2000, 21: 1104～1115

37 Cordwell S J, Nouwens A S, Verrils N M, Basseal D J, Walsh B J. Sub-proteomics based upon protein cellular location and relative solubilities in conjunction with composite two-dimensional gels. Electrophoresis, 2000, 21: 1094～1103

38 Molloy M P, Herbert B, Walsh B J, Tyler M I, Traini M, Sanchez J C, Hoehstrasser D F, Williams K L, Gooley A A. Extraction of membrane proteins by differential solubilization for separation using two-dimensional electrophoresis. Electrophoresis, 1998, 19: 837～844

39 Rabilloud T, Adessi C, Giraudel A, Lunardi J. Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis, 1997, 18 (3-4): 307～316

40 Mark P, Molloy, Ben R, Herbert, et al. Extraction of membrane proteins by differential solubilization for separation using two-dimensional gel electrophoresis. Electrophoresis, 1998, 19: 837～844

41 Rabilloud T, Blisnick T, Heller M, Luche S, Aebersold R, Lunardi J, Braun-Breton C. Analysis of membrane proteins by two-dimensional electrophoresis: comparison of the proteins extracted from normal or Plasmodium falciparum-infected erythrocyte ghosts. Electrophoresis, 1999, 20 (18): 3603～3610

42 Cordwell S J, Nouwens A S, Verrills N M, McPherson J C, Hains P G, Van Dyk D D, Walsh B J. The microbial proteome database—an automated laboratory catalogue for monitoring protein expression in bacteria. Electrophoresis, 1999, 20: 3580～3588

43 Gerner C, Holzmann K, Meissner M, Gotzmann J, Grimm R, Sauermann G. Reassembling proteins and chaperones in human nuclear matrix protein fractions. J Cell Biochem, 1999, 74: 145～151

44 Kitakawa M, Graack H R, Grohmann L, Goldschmidt-Reisin S, Herfurth E, Wittmann-Liebold B, Nishimura T, Isono K. Identification and characterization of the genes for mitochondrial ribosomal proteins of Saccharomyces cerevisiae. Eur J Biochem, 1997, 245: 449～456

45 Tsiboli P, Herfurth E, Choli T. Purification and characterization of the 30S ribosomal proteins from the bacterium Thermus thermophilus. Eur J Biochem, 1994, 226: 169～177

46 Carroll J A, Garon C F, Schwan T G. Effects of environmental pH on membrane proteins in Borrelia burgdorferi. Infect Immun, 1999, 67: 3181～3187

47 Moroni A, Pavan G, Donati M, Cevenini R. Differences in the envelope proteins of Chlamydia pneumoniae, Chlamydia trachomatis, and Chlamydia psittaci shown by two-dimensional gel electrophoresis. Arch Microbiol, 1996, 165: 164～168

48 Tsugita A, Kawakami T, Uchiyama Y, Kamo M, Miyatake N, Nozu Y. Separation and characterization of rice proteins. Electrophoresis, 1994, 15: 708～720

49 Sonnenberg M G, Belisle J T. Definition of Mycobacterium tuberculosis culture filtrate proteins by two-dimensional polyacrylamide gel electrophoresis, N-terminal amino acid sequencing, and electrospray mass spectrometry. Infect Immun, 199?, 65: 4515－4524

50 Wallin E, von Heijne G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Prot Sci, 1998, 7: 1029～1038

51 Paulsen I T, Sliwinski M K, Nelissen B, Goffeau A, Saier M H Jr. Unified inventory of established and putative transporters encoded within the complete genome of Saccharomyces cerevisiae. FEBS Lett, 1998, 430: 116～125

52 De Michelis M I, Spanswick R M. H^+-pumping driven by the varadate-sensitive ATPase in membrane vesides from roots. Plant Physiol, 1986, 81: 542～547

53 Bordier C. Phase separation of integral membrane proteins in Triton X-114 solution. J Biol Chem, 1981, 256: 1604～1607

54 Chevallet M, Santoni V, Poinas A, Rouquie D, Fuchs A, Kieffer S, Rossignol M, Lunardi J, Garin J, Rabilloud T. New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis. Electrophoresis, 1998, 19: 1901～1909

55 Santoni V, Kieffer S, Desclaux D, Masson F, Rabilloud T. Membrane proteomics: use of additive main effects with multiplicative interaction model to calssify plasma membrane proteins according to their solubility and electrophoretic properties. Electrophoresis, 2000, 21: 3329－3344

56 Santoni V, Molloy M, Rabilloud T. Membrane proteins and proteomics: un amour impossible? Electrophoresis, 2000, 21: 1054－1070

57 Molloy M P, Herbert B R, Walsh B J, Tyler M I, Traini M, Sanchez J C, Hochstrasser D F, Williams K L, Gooley A A. Extraction of membrane proteins by differential solubilization for separation using two-dimensional gel electrophoresis. Electrophoresis. 1998, 19: 837－844

58 Gyenes T, Gyenes E. Effect of "stacking" on the resolving power of ultrathin-layer two-dimensional gel electrophoresis. Anal Biochem, 1987, 165: 155～160

59 Rabilloud T, Gianazza E, Catto N, Righetti P G. Amidosulfobetaines, a family of detergents with improved solubilization properties: application for isoelectric focusing under denaturing conditions. Anal Biochem, 1990, 185: 94～102

60 Muller E C, Schumann M, Rickers A, Bommert K, Wittmann-Liebold B, Otto A. Study of Burkitt lymphoma cell line proteins by high resolution two-dimensional gel electrophoresis and nanoelectrospray mass spectrometry. Electrophoresis, 1999, 20: 320－330

61 Jung E, Hoogland C, Chiappe D, Sanchez J C, Hochstrasser D F. The establishment of a human liver nuclei two-dimensional electrophoresis reference map. Electrophoresis, 2000, 21: 3483－3487

62 Jung E, Heller M, Sanchez J C, Hochstrasser D F. Proteomics meets cell biology: the establishment of sub-

cellular proteomes. Electrophoresis, 2000, 21: 3369－3377

63 Wirth P J, Luo L D, Benjamin T, Hoang T N, Olson A D, Parmelee D C. The rat liver epithelial (RLE) cell nuclear protein database. Electrophoresis, 1993, 14: 1199－1215

64 Fey E G, Penman S. Nuclear matrix proteins reflect cell type of origin in cultured human cells. Proc Natl Acad Sci USA, 1988, 85: 121

65 Gerner C, Sauermann G. Nuclear matrix proteins specific for subtypes of human hematopoietic cells. J cell Biochem, 1999, 72: 470－482

66 Otto H, Dreger M, Bengtsson L, Hucho F. Identification of tyrosine-phophorylated proteins associated with the nuclear envelope. Eur J Biochem, 2001, 268: 420－428

67 Dreger M, Bengtsson L, Schoneberg T, Otto H, Hucho F. Nuclear envelope proteomics: novel integral membrane proteins of the inner nuclear membrane. Proc Natl Acad Sci USA. 2001, 98 (21): 11943－11956

（万晶宏）

第四章　二维电泳与细胞蛋白质的分离

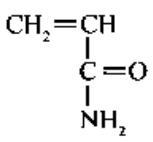
第一节　概　　述

生物大分子如蛋白质、核酸、多糖等常以颗粒分散在溶液中，它们所带电荷的性质取决于溶液的酸碱度以及其他相互作用的因素如十二烷基硫酸钠所引起的酸碱度的变化。在电场中，带电离子向正极或负极移动，迁移的方向和速度取决于这些离子所带的电荷和介质的阻力，这种迁移现象即称为电泳。迁移的原理可分为3类：根据分子大小和形状、分子所带电荷、或分子的其他生物学及物理化学特性。为减小电泳过程中由于发热和扩散效应所引起的分子区带扭曲和扩散，目前的电泳大多是在固定化的介质中将样品放入流动相中进行分离，而不采用完全自由态的溶液进行分离。根据所用介质种类不同，电泳可分为两类：纸、乙酸纤维素、薄层材料如硅胶、氧化铝、纤维素等惰性程度较高，主要用于支持体系，且变形不大，用这些介质分离蛋白质，主要依赖于蛋白质的电荷密度；琼脂糖、淀粉及聚丙烯酰胺凝胶体系可变性较强且有一定的扩散作用，但这些体系在分离蛋白质时，除依赖于蛋白质的电荷密度外，还与蛋白质的相对分子质量大小及分子尺寸相关，提高了对蛋白质的分辨能力。由于聚丙烯酰胺凝胶的凝胶孔径与蛋白质分子大小接近，能达到最好的分离能力，而且聚丙烯酰胺凝胶制胶方便，代价低，重现性好，凝胶透明，易于显色并观察，此外与下游蛋白质鉴定方法如免疫印迹、质谱鉴定等兼容性较好，因而使得聚丙烯酰胺凝胶成为首选的蛋白质分离介质[1,2]。

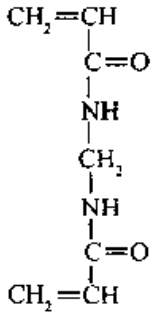
聚丙烯酰胺凝胶（polyacrylamide gel，PAG）是通过丙烯酰胺（acrylamide）单体和双功能团交联剂如 N，N'-甲叉双丙烯酰胺（N，N'-methylene bisacrylamide，常称甲叉丙烯酰胺）按一定比例混合，在有引发剂和增速剂存在下聚合而形成的交叉网状结构，丙烯酰胺单体聚合生成长链聚合物，甲叉丙烯酰胺的双功能基团和链末端的自由功能基团反应从而发生交联，形成网状结构。两个单体的常用物化属性见表4-1，两个单体和聚丙烯酰胺凝胶的交联结构见图4-1[3]。根据分离目的不同，可以按不同浓度及比例来配制PAG胶。如：等电聚焦凝胶只需按等电点不同分离蛋白质，而对凝胶按相对分子质量大小的筛分能力要求不高，用低浓度胶即可；在利用十二烷基硫酸钠-聚丙烯酰胺凝胶电泳（SDS-PAGE）分离蛋白质时，不仅要根据蛋白质的带电密度分离，而且还要根据被分离样品的相对分子质量大小设计相应浓度的PAG胶。不同浓度的凝胶，其机械

表 4-1　丙烯酰胺凝胶两个单体的常用物化属性

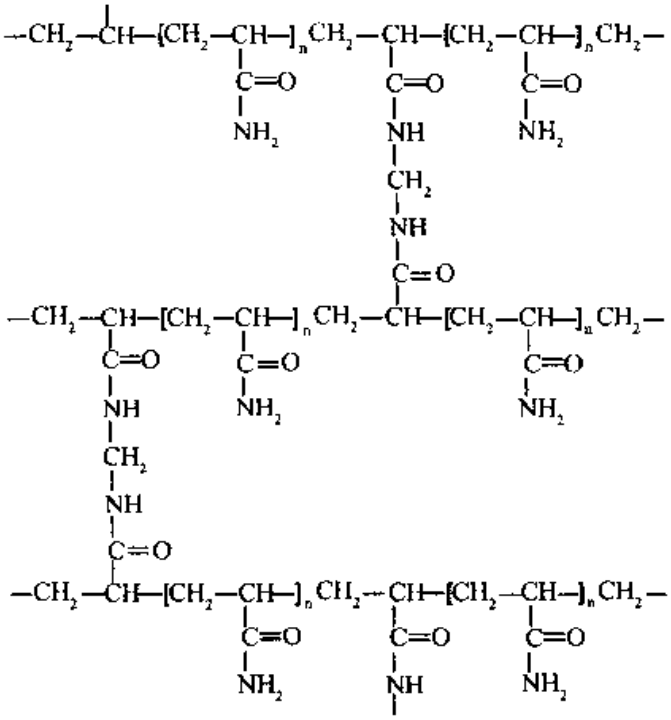
属性	丙烯酰胺	*N*，*N*′-甲叉双丙烯酰胺
分子式	$CH_2{=}CH{-}CO{-}NH_2$	$(CH_2{=}CH{-}CO{-}NH)_2CH_2$
分子质量	71.08	154.17
外观	白色结晶粉末	白色结晶粉末
气味	无	无
贮存	密封、避光	冷、密封、避光
毒性	中枢神经毒物	中枢神经毒物



丙烯酰胺



甲叉丙烯酰胺



聚丙烯酰胺凝胶

图 4-1　构成聚丙烯酰胺凝胶的两个单体的结构及聚合物的网状结构示意图

性能、弹性、透明度、黏性及孔径大小均不同，这些性能取决于凝胶配制时的两个重要参数 T 和 C[4]。T 是两个单体（丙烯酰胺和甲叉丙烯酰胺）的总百分浓度，C 是甲叉丙烯酰胺占两个单体总重的百分含量。具体计算公式如下：

$$T = \frac{a + b}{m} \times 100 \quad (\%)$$

$$C = \frac{b}{a + b} \times 100 \quad (\%)$$

这里 a 为丙烯酰胺的质量（g），b 为甲叉丙烯酰胺的质量（g），m 为凝胶聚合前的溶液体积（ml）。T 和 C 的值是很关键的。一般的蛋白分离胶 T 值分布在5%到30%之间，T 值太小凝胶黏度增加，难以处理，太大则凝胶脆性增加，容易破裂。C 值一般在3%左右，适合于大多数的蛋白质分离。根据我们的经验，第二向电泳要分离蛋白混合物，固定 C 值为2.6%（0.8/30），预实验时取 T 值为13%，随后根据目的蛋白的相对分子质量来调整 T 值，会获得比较满意的结果。

Smithies 和 Poulik[5]（1956）最早引入了二维电泳技术，他们将纸电泳和淀粉凝胶电泳结合来分离血清蛋白质。随后二维电泳技术经历了更多的发展[6]，并将聚丙烯酰胺介质引入应用，特别是等电聚焦技术在一向的应用，使基于蛋白质电荷属性的一向分离成为可能。目前所应用的二维电泳体系是由 O'Farrell 等首先于1975年发明[7]，其原理是根据蛋白质的两个一级属性：等电点和相对分子质量的特异性，将蛋白质混合物在第一个方向上按照等电点高低进行分离（等电聚焦），在第二个方向上按照相对分子质量大小进行分离（十二烷基硫酸钠-聚丙烯酰胺凝胶电泳，SDS-PAGE）（图4-2）。在最好状态下，传统等电聚焦技术和 SDS-PAGE 在 20 cm 左右长度的凝胶中可在各自方向上分辨出100个不同的蛋白质条带，因此，理论上的二维电泳分辨能力可达到 10 000 个点。目前已有实验室在 30 cm×40 cm 的大胶上获得这一分离效果。考虑到大多数实验室并不具备运行这种大胶所需要的条件，普通情况下的凝胶（20 cm×20 cm）分辨到3000个点已是相当不错。但利用特殊的样品制备方法如三步提取或亚细胞器提取，以及下文将要提到的窄范围等电聚焦胶条，都可大大提高二维电泳的分辨能力。二维电泳分离后的蛋白质点经显色方法如考马斯亮蓝染色、酸性银染、碱性银染、负性染色、荧光染色或放射性标记等染色处理后，通过图像扫描存档，最后呈现出来的是在二维方向排列的呈“满天星”状排列的小圆点，其中每一个点代表一个蛋白质。

图4-3是我们分离得到的 HL-60 细胞裂解蛋白质的二维电泳图谱，从该图谱可以初步得到该细胞内蛋白质的分布范围（偏酸性或偏碱性）、表达蛋白质的个数（凝胶点的数目）、蛋白质的大概相对分子质量、等电点（通过相对分子质量和等电点分析校正）、蛋白质相对表达丰度（点的灰度值）。需要说明的是，由于

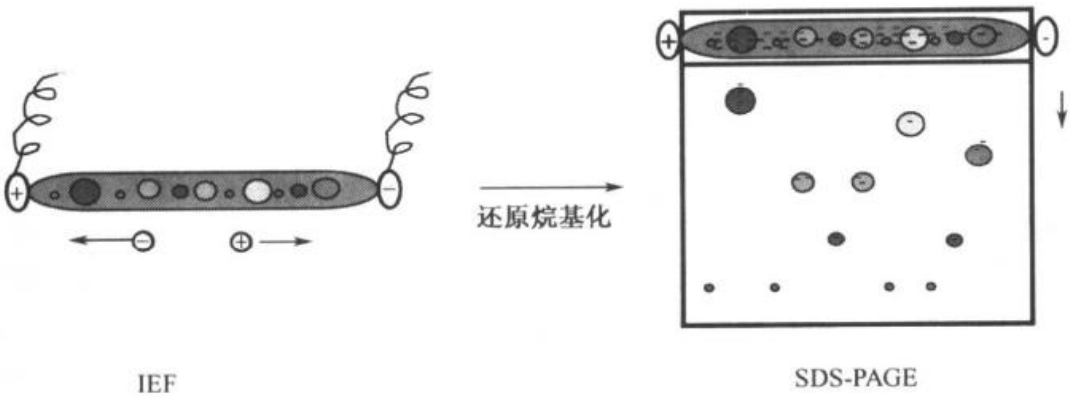


图 4-2　二维电泳示意图

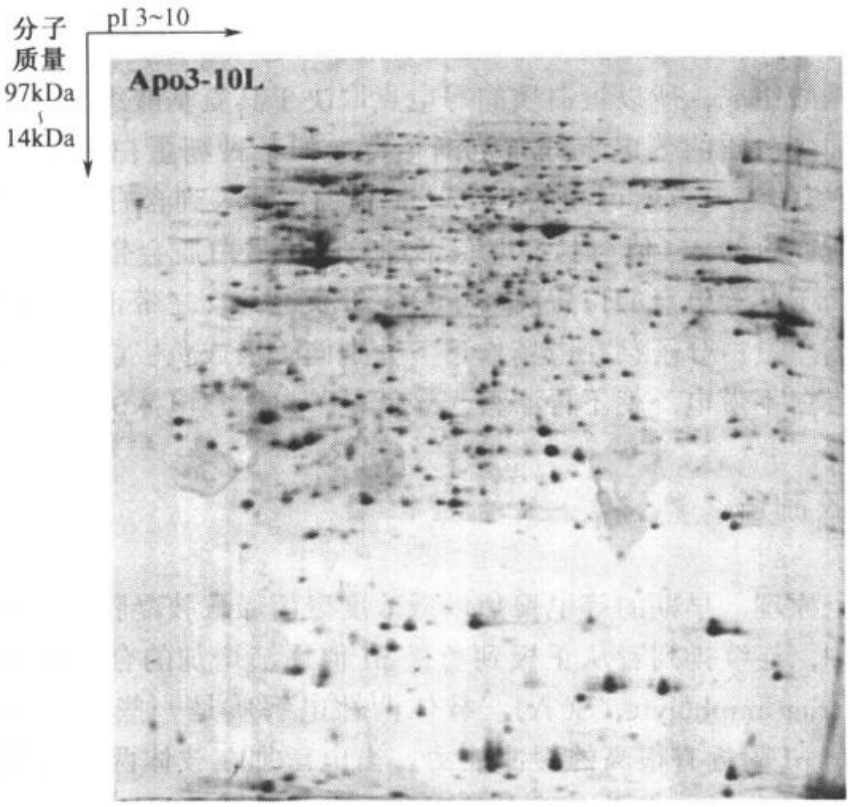


图 4-3　ATRA 诱导 HL-60 细胞凋亡启动时的蛋白二维凝胶电泳
（银染扫描图）
100 μg HL-60 全细胞裂解蛋白，利用分析型二维电泳，在 pH3～10 的范围内进行等电聚焦和 14% 聚丙烯酰胺凝胶浓度进行聚丙烯酰胺凝胶电泳，银染的凝胶经图像分析。

在电泳过程中涉及到亚基内或亚基间二硫键的还原和烷基化处理，亦即高级结构的去除，因此，通过二维电泳分离所得到的实质上是构成蛋白质的各个亚基，而

非完整的功能蛋白质。

第二节　一维等电聚焦电泳

蛋白质是由 20 种不同氨基酸按不同比例通过肽键的连接构成的。由于构成蛋白质的一些氨基酸侧链在一定的 pH 值的溶液中是可解离的，从而可带有一定的电荷。如酸性氨基酸天冬氨酸和谷氨酸在酸性条件下，其末端羧基将只带一个质子而呈中性状态，但在碱性条件下，末端羧基将失去一个质子而带负电。而碱性氨基酸精氨酸、赖氨酸的氨基和组氨酸的咪唑基在酸性条件下则结合质子而带正电，碱性条件下，这些质子解离，残基呈中性。构成蛋白质的所有氨基酸残基上所带正负电荷的总和便是蛋白质所带的静电荷。在低 pH 条件下，蛋白质的静电荷为正；而在高 pH 条件下，其静电荷为负。若在某一 pH 值的溶液中，蛋白质的静电荷为零，则此 pH 值即为该蛋白质的等电点（pI）。由于不同蛋白质有着不同的氨基酸组成，所以蛋白质的等电点取决于其氨基酸组成，是一个物理化学常数。不同蛋白质的等电点分布范围很宽，如某种糖蛋白的等电点值可低至 1.8，而溶菌酶的等电点值可高达 11.7。把蛋白质加入到含有 pH 梯度的载体时，如果蛋白质所在点的 pH 值与其等电点不符，则该蛋白质会带一定量的正电荷或负电荷。在高于其等电点的位置时，蛋白质带负电，反之带正电。此时，如果外加一个强电场，蛋白分子会在电场作用下分别向正极或负极漂移，当达到其等电点位置时，蛋白不带电，就不再漂移，这就是等电聚焦电泳的基本原理。

一、IEF 凝胶制备

根据以上原理，早期的等电聚焦以低浓度聚丙烯酰胺凝胶为介质，在外加电场的该介质中，连续排列着从正极到负极 pI 值逐渐增加的合成载体两性电解质（synthetic carrier ampholyte，SCA）。载体两性电解质是一些可溶性的两性小分子，它们在其 pI 附近有很高的缓冲能力。当电压加在载体两性电解质混合物间时，最高 pI 值的分子（带最多正电荷）移向负极，最低 pI 值（带最多负电荷）的分子移向正极，其余分子将根据其 pI 值在两个极值之间分散，从而形成一个连续的 pH 梯度。根据等电聚焦的基本原理，当蛋白混合物加入到固相介质中并处于外加电场下，各个蛋白质将在迁移至与其等电点相同的 pH 位置时停止移动，一旦偏离该位置，该蛋白质将带与其偏离方向相同的电荷，因而向另外一个方向移动，直至达到动态平衡，从而实现根据等电点不同而在电荷方向的分离。载体两性电解质的引入是等电聚焦技术的一大突破，但该技术仍然存在以下几点缺陷：

1）由于 SCA 是通过复杂的合成过程得到的，其重复性很难控制，由此不同

制备批次之间会存在相当大的变化，从而使一向分离时，同一蛋白质在不同批次等电聚焦中所出现的位置有所偏差。这样就限制了 2-DE 分离的重复性。

2) SCA 分子相对较小，难以在 IEF 胶内固定，在 IEF 过程中由水合正离子引起的电渗流（electroendosmosis）将致使 SCA 分子向负极迁移（负极漂移），结果会使 pH 值不稳定性增加。当利用管状 IEF 胶时，由于玻璃毛细管壁表面的硅羟基带负电，与这些负电荷基团对应，将在管胶表面聚集一层正电荷形成双电层，这种作用加剧了负极漂移。

3) 由于负极漂移作用对碱性区蛋白质的影响尤其严重，结果常会导致碱性区蛋白质难以成功聚焦甚至导致碱性区蛋白质的丢失。

4) 每一次灌制 SCA 凝胶的重复性难以控制，而且这种凝胶机械稳定性差，易拉伸变形或断裂，同样导致重复性的降低。

由 O'Farrell 所推出的非平衡 pH 梯度电泳（non-equilibrium pH gradient electrophoresis, NEPHGE），在快速形成的 pH 梯度中根据蛋白质移动能力进行分离，该法可以用来分离碱性蛋白质[8]，但在结果的重现性上，仍然没有很好的解决。真正的一向等电聚焦技术的突破要归功于在 Immobilines 试剂（Amersham Pharmacia Biotech, APB）的基础上开发的固相 pH 梯度（immobilized pH gradient, IPG）技术[9]。Immobilines 是拥有 CH_2═CH—CO—NH—R 结构的 8 种丙烯酰胺衍生物系列，其中 R 包含羧基或叔氨基团，它们构成了分布在 pH3～10 范围不同值的缓冲体系。根据一定的配方计算后，将适宜的 IPG 试剂添加至混合物中用于凝胶聚合，在聚合中，缓冲基团通过乙烯键共价聚合至聚丙烯酰胺骨架中，形成 pH 梯度。为了使操作方便，IPG 胶在一层塑料支持膜上聚合，在洗去引发剂和剩余单体后，将干燥后的胶板切成 3mm 宽的干胶条，保存于 −20°C 备用。与传统载体两性电解质预制胶相比，IPG 胶具有机械性能好、重现性好、易处理、上样量大的特点，而且避免了电渗透作用，因而可以进行特别稳定的 IEF 分离，达到真正的平衡状态。由于 Immobilines 可以随意调配，这样就可以控制生产不同 pH 分离范围的干胶条。如宽范围胶条有 pH3～10、pH3～12，窄范围胶条有 pH4～7、pH6～11、pH5～8 等，甚至可以限定到一个 pH 单位，可以根据样品的分布范围来选择胶条 pH 值范围。Amersham Pharmacia Biotech 公司和 Bio-Rad 公司都有大量不同 pH 范围的商业化胶条可供选择（见表 4-2）。此外根据需求，还可以选择不同长度的 IPG 胶条，如 APB 公司目前就有 7 cm、11 cm、13 cm、18 cm、24 cm 五种长度的胶条可供选择。专门配合 IPG 胶的等电聚焦设备已商品化，如 APB 公司的 IPGphor 及其附件，Bio-Rad 公司的 PROTEAN IEF（图 4-4），其编程自动化及高至 8000V 或 10 000V 的电压使得 IEF 在一个晚上即可完成，大大缩短了 2-DE 的进程。虽然早期 IPG 技术在 2-DE 应用中遇到很多问题，但它们被逐渐克服[10～12]并使得 IPG IEF 成为目前大多数 2-DE 分离首选

的方法。这种 2-DE 系统也被称为 IPG-DALT 系统，本章所讨论的二维电泳流程也是按此系统来阐述。

表 4-2　部分可选的窄范围 IPG 干胶条

公司	Amersham Bioscience	Bio-Rad
胶条长度（cm）	7、11、13、18、24	7、11、17
pH 范围	3.5～4.5、4.0～5.0、4.5～5.5、5.0～6.0	3～6
	5.5～6.7、6～9、3～7 NL、3～10、3～10 NL	5～8
	4～7	7～10
	（胶条长度为 18 cm 或 24 cm）	3～10

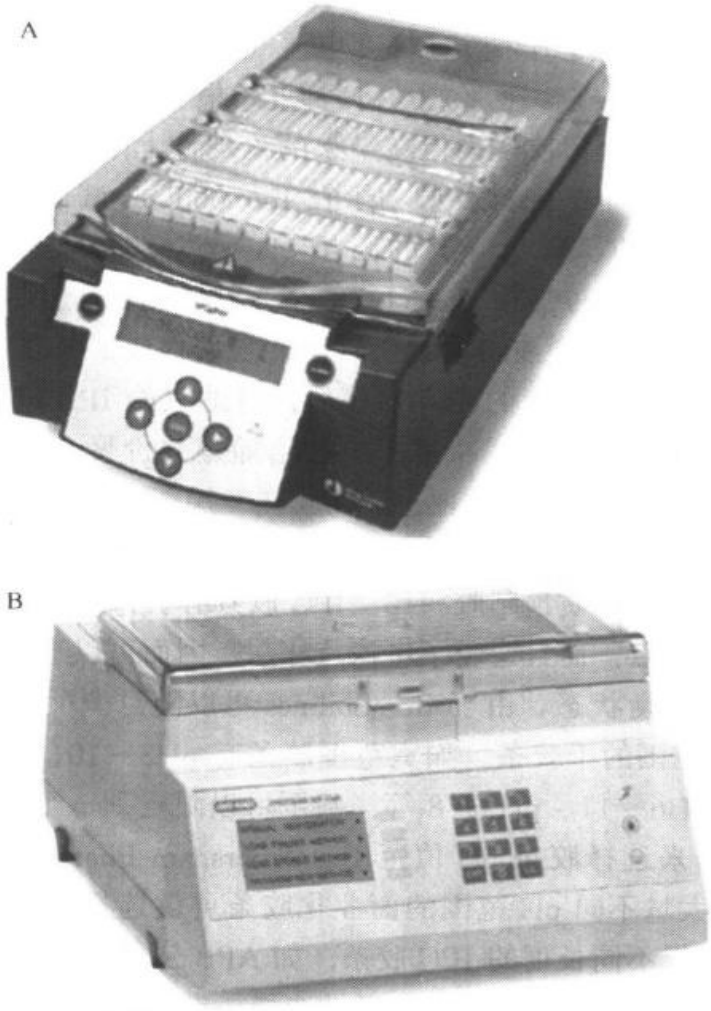


图 4-4　等电聚焦仪

A. APB 公司 IPGphor。可以同时运行 12 根胶条，自动控制，自动冷却，电压可升至 8000V，可设定九个不同程序。B. Bio-Rad 公司 PROTEAN IEF。可以同时运行 12 根胶条，电压可升至 10 000V。

二、加样

虽然前面的章节已经详细讲解了关于样品制备的各个方面及应注意事项，在谈到加样前还是首先要强调一下样品制备。因为对于一个熟练的二维电泳操作人员来说，从加样的那一刻起，基本上已决定了这一周期 2-DE 实验的结果成功与否。特别是对较为特殊的样品，如亚细胞器裂解液、膜蛋白提取物等，必要时需采用特殊助溶试剂或三步提取等制备方法（参考第三章）。针对不同组织、细胞、体液、植物或全生物体应有相应的样品制备方法，常用方法有匀浆（粉碎）-三氯乙酸/丙酮沉淀-干燥法、细胞的裂解液提取法、体液（血清）的重泡胀液稀释直接上样法等。因此，对于首次接触到制备样品或二维电泳的人来说，利用小胶（如 7 cm 的 IPG 胶条）做一次预实验，可以节约时间，观察样品制备的好坏，积累经验，以避免不必要的损失。

对一新样品，常使用宽范围、线性 pH3.5～10 的 IPG 梯度。但对大多数样品，这样做会丧失 pH4～7 区域的分辨率，许多蛋白质的 pI 值在该区域。利用非线性 pH3.5～10 IPG IEF 胶在一定程度上缓解了这个问题[13]。pH3.5～10 非线性胶条在 pH4～7 区域的梯度要比 pH7～10 更为平坦，保证在大部分碱性蛋白质都能得到分辨的前提下，pH4～7 区域能得到更好的分离。但用 pH4～7 的 IPG IEF 胶则能在该范围得到更好的分离效果（图 4-5）。而且使用窄范围 IPG 胶的另一个优点是可以加大上样量，这样可以观察到更多的蛋白质点。这些 pH 范围的胶条目前已经商品化（Amersham Pharmacia Biotech，Bio-Rad）。实验室自制 pH4～9 的 IPG 胶也能覆盖来自许多不同种类样品的大多数蛋白质。这些梯度的 IPG 胶与加样杯上样或胶内泡胀兼容，在 Multiphor 和 IPGphor 系统都工作良好，均适用于分析胶（上样量 50～100 μg）或微量制备胶（上样量达 1 mg）。加样方案有两种，一种是 IPG 胶重泡胀后，利用加样杯，边运行等电聚焦，边上样，利用加样杯上样的好处在于加样量可以提高很多，由于加样杯直接和胶面接触，使得分子质量大于 100000Da 的蛋白质可以有效的进入 IPG 胶条；另一种方案是将样品与重泡胀液混合，在 IPG 胶条泡胀的同时样品也渗入胶条，然后再加电压，该方法被称为胶内泡胀法，其优点是泡胀和等电聚焦整合为一个程序即可完成，提高了工作效率，保证了重复性，这是目前常用的加样方法。样品通常是与重泡胀液混合后加在持胶槽（strip holder）的电极内侧，然后选择所需 pH 范围的胶条放于持胶槽中，将样品溶液从正极到负级均匀铺展。对不同类型的样品，最好的加样位置由经验值确定，通常情况下加在正极端。常用的重泡胀液成分为 8 mol/L 尿素的溶液（或 2 mol/L 硫脲和 7 mol/L 尿素），其他成分为 2% 非离子去污剂（NP-40，Triton X-100）或两性去污剂3-[(3-胆酰胺丙基)-二乙胺]-丙磺酸(3-[(-cholamidopropyl)-dimethylammonio]-1-propane sulphonate, CHAPS),

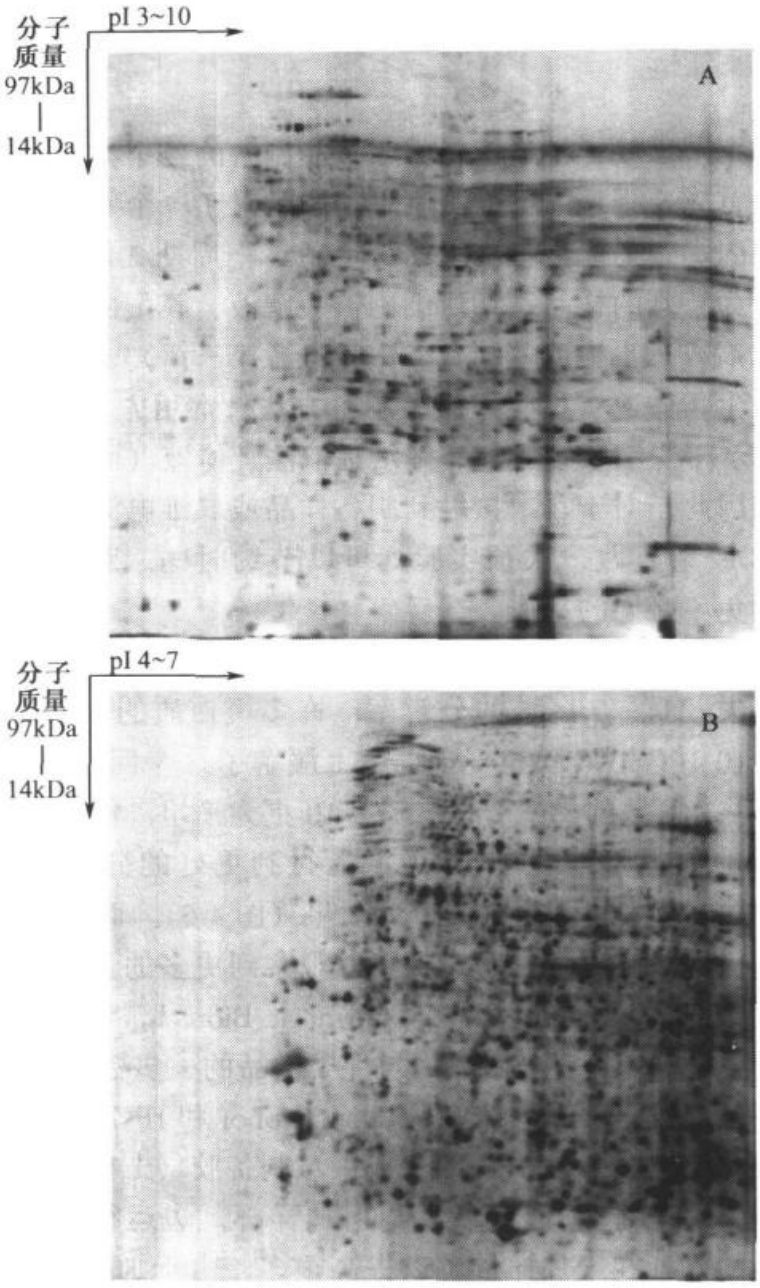


图 4-5　胎肝组织提取物不同 pH 范围
IPG-DALT 二维电泳分离图谱

A. 一维：IPG3～10 非线性（预制胶）。分离距离 18 cm。胶内泡胀法上样（200 μg）。二维：垂直 SDS-PAGE（13%T，连续）。银染。B. 一维：IPG4～7（预制胶）。分离距离 18 cm。胶内泡胀法上样（1000 μg）。二维：垂直 SDS-PAGE（12%T，连续）。银染。

15 mmol/L DTT 和 0.5%适宜范围 pH 值的 IPG 缓冲液（IPG buffer）。由于尿素的浓度很高，样品重泡胀过程中水的挥发可能导致尿素结晶，影响胶条泡胀和 IEF 时样品的顺利迁移，因此要在加样后的胶条上覆盖一层惰性矿物油，可以阻

止水分挥发，同时也保护样品以避免高压 IEF 过程中的氧化作用。重泡胀过程通常不加电压，一般要持续 10～12 h，但也有研究发现在重泡胀过程中加 30V 的低电压，可以使样品聚焦更好。上样量要根据显色方法和需求而定，如果是想观察一个蛋白质表达全谱，可以用银染做分析胶，这时对于 18 cm 的 IPG 胶条，上样量在 100～150 μg 为宜。而如果是要进行考马斯亮蓝染色或质谱分析、转膜分析，此时的制备胶上样量可以从 500 μg 到几个毫克。

三、运行

重泡胀完成后，可以加电压开始 IEF。由于胶条所能承受的电流有限，因此要避免电流过大，一般在 IEF 过程中限流为 50 μA。最初时由于样品中离子强度高，电压应限制于 200V、30 min，然后增加电压至 500 伏运行 1 h，并逐渐增加电压至 1000V，最终电压至 8000V 并恒定运行若干小时。运行时间决定于几个不同的因素，包括样品类型、蛋白上样量、IPG 胶条长度及所用 pH 梯度。理论上讲，获得最好的图谱质量和重复性所需最佳时间是 IEF 分离达到稳定态所需的时间[10,14,15]。若聚焦时间太短，会导致水平和垂直条纹，但要避免过度聚焦。虽然与经典 O'Farrell 法相比，不会导致蛋白质向负极的迁移（负极漂移），但却会因为活性水转运而导致过多水在 IPG 胶表面渗出（电渗），会造成蛋白图谱变形。在胶条碱性端产生水平条纹以及蛋白质丢失。因为温度低时尿素可能会结晶，IEF 应当在 20℃运行，不同的温度下在 2-DE 图谱中一些蛋白质点的相对位置会发生变化[16]。已经商品化的仪器如 APB 公司的 IPGphor 和 Bio-Rad 公司的 PROTEAN IEF 都能设定程序自动完成这些操作，而且能同时运行 12 根胶条，而不像早期使用 Multiphor II（APB 公司）水平多用电泳仪时需要在重泡胀后重新用加样杯上样。针对不同 pH 范围的 IPG 胶条，表 4-3 中列出了在使用 IPGphor 时一些典型的运行条件以便参考。整个 IPG IEF 的流程如图 4-6 所示。

表 4-3　利用 IPGphor 的运行条件[11]

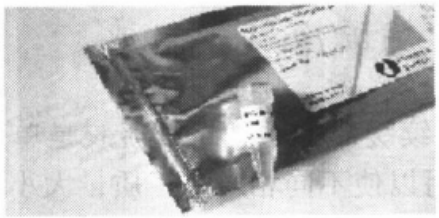
胶条长度	180 mm
温度	20℃
最大电流	每个胶条 50 μA
最大电压	8000V
分析型 IEF	
重泡胀	◆ 30V，12～16h
初始化 IEF	◆ 200V，1h
	◆ 500V，1h
	◆ 1000V，1h

续表

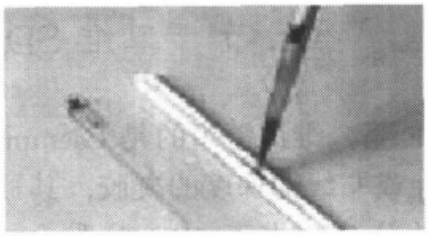
达到稳定态后 IEF	◆ 30 min 内由 1000V 梯度上升至 8000V
	◆ 8000V 达到稳定态，根据所用胶条 pH 间隔而定
1～1.5 个 pH 单位	IPG5～6　8h
	IPG4～5.5　8h
3 个 pH 单位	IPG4～7　4h
4 个 pH 单位	IPG4～8　4h
5～6 个 pH 单位	IPG4～9　4h
Ⅱ 微量制备 IEF	
重泡胀	30V，12～16h
达到稳定态后 IEF	相应分析胶 IEF 时间延长约 50%

四、IPG 胶条的平衡

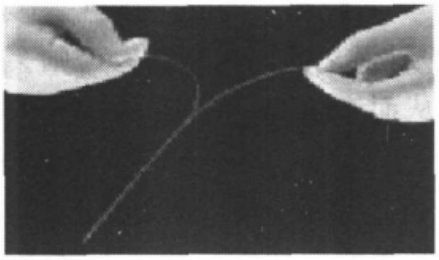
在一向 IPG IEF 后，胶条可马上用于二向，也可保存在两片塑料膜间于－80℃存放几个月时间。在二向分离前，必须要平衡 IEF 胶。平衡过程需两个步骤，第一步的平衡液成分为 50 mmol/L Tris 缓冲液（pH8.8），含 2% W/V SDS、20 mmol/L DTT 6 mol/L 尿素和 30%甘油。平衡液的主要作用是使一向胶条上的蛋白质变性。SDS 是一种阴离子去污剂，在溶液中当 SDS 单体浓度大于 1 mmol/L时，可以和蛋白定量结合，蛋白质与 SDS 重量比为 1∶1.4，此时蛋白质所带负电荷过量，而且与蛋白质相对分子质量成正比关系。在电场作用下，各蛋白质依所带负电荷数目迁移，迁移速率与电荷数成正比关系，亦即实现了蛋白质相对分子质量方向上的分离。为了让 SDS 与蛋白质充分结合，必须加入尿素和还原剂等，以去除蛋白的高级结构以及亚基之间相互作用，尿素和甘油同时也用于减缓电渗效应，否则会使一向到二向的蛋白质转移率降低[10]。加入还原剂 DTT 是为了在蛋白质与 SDS 充分结合的同时，二硫键也得到还原。整个第一步大约需持续孵育 15 min。接下来同样是在缓冲液中平衡 15 min，只不过此时用 100 mmol/L 碘乙酰胺取代 DTT，这一步是用来烷基化自由 DTT 和蛋白质所带的自由疏基，否则自由 DTT 在二向 SDS-PAGE 胶迁移，会产生点条纹假像，在银染后会观察出来。另外一种可以一步完成平衡的方法是在平衡液中用 5 mmol/L TBP 取代 DTT，TBP 不带电荷，在 SDS-PAGE 过程中不迁移[17]。平衡后上二向胶之前，IPG 胶应沿边缘用滤纸吸 1 min 以去除多余平衡液。



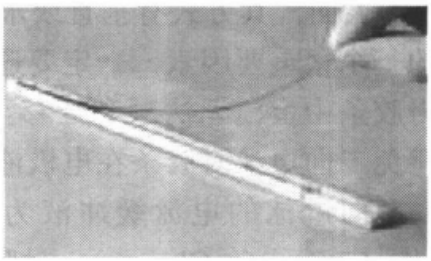
A 胶条和 IPG 缓冲液



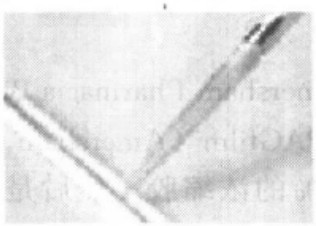
B 加入样品和重泡胀液



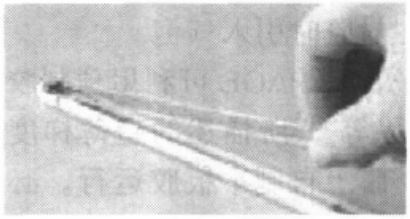
C 去除 IPG 胶保护膜



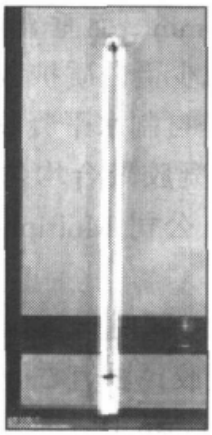
D 放入胶条，铺展样品



E 用覆盖油封胶



F 封闭持胶槽



G 转移持胶槽

H 编程并运行

图 4-6　IPG IEF 流程

一向预制胶条、IPG 缓冲液和设备为 APB 公司产品。上样方法采用胶内泡胀法。

第三节　二维 SDS-聚丙烯酰胺凝胶电泳

2-DE 第二维最常用的是 Laemmli[18] 非连续缓冲体系，二维凝胶是单一浓度或含线性或非线性梯度的凝胶，其覆盖范围可以使不同相对分子质量大小蛋白质得到有效分离。由于 IPG 胶黏合在弹性塑料支持膜上，处理起来比用经典 O'Farrell 体系中较脆的 IEF 管胶容易。平衡后，IPG 胶条可直接加在二向 SDS-PAGE 胶的表面，其方式有垂直或水平两种。一向胶和二向胶的接触是影响电泳重复性的一个很重要因素，一定要避免两者接触面之间产生气泡，否则会产生阻力，使得胶条中蛋白无法顺利迁移至二向，从而产生点的扭曲现象。在垂直胶中，为避免二向电泳时胶条在电极液中移位，需用 0.5% 的琼脂糖电极缓冲液溶液封胶，二向电泳的电泳缓冲液为 Tris-Glycine-SDS 系统（500ml，5×，7.5g Tris，2.5g SDS，36g Glycine）。封胶时注意琼脂糖溶液的温度不能太高，否则会造成 IPG 胶上的蛋白质变性或产生蛋白修饰。因为琼脂糖封胶液凝固很快，要避免在封胶时引入气泡。

水平 SDS-PAGE 可利用预制胶（ExcelGel，Amersham Pharmacia Biotech）或依据 Görg 等[10] 所描述的超薄梯度胶在 GelBond PAGfilm（Amersham Pharmacia Biotech）膜上自制平板胶运行。必须要用 T 为 6% 的浓缩胶，然后是均一（如 12%）或梯度（如 12%～15%）分离胶，该方法在 Görg 等[11] 中有详细描述。水平二向系统的优点是凝胶附着在塑料支持膜上，在染色过程中可以防止凝胶大小发生变化。而且由于水平胶厚度较小（一般在 0.5 mm，而垂直胶为1 mm或 1.5 mm），因而可以施加较高电压，减短运行时间，减少蛋白质扩散，使得水平胶分离蛋白质点的边缘要比垂直胶清晰。此外，如果是自制水平凝胶，由于水平胶灌制时边缘密封作用好，使水平胶与垂直胶相比，凝胶聚合均匀，边缘效应小，蛋白质点的分布变形程度小。图 4-7（A）为 APB 公司 Multiphor II 可支持水平胶运行的多用电泳仪。

大规模蛋白质组分析通常需要成批二向 SDS-PAGE 同时运行以获得最好的图谱重复性。而水平电泳则不能同时跑两块较大的凝胶，垂直二向 SDS-PAGE 则具有这方面的优点，垂直电泳仪如 APB 公司的 Ettan Dalt II［图 4-7（B）］，可以同时运行 12 块二向胶。与水平胶相比，垂直胶具有操作简单、上样量大、重复性好等优点。在垂直电泳系统中不需要浓缩胶，因为在 IPG 胶条中蛋白质区已得到浓缩[19]。除 Ettan 系列外，多块胶垂直二维 SDS-PAGE 还有 Investigator（Genomic Solutions）以及 Protean II xi（Bio-Rad）［图 4-7（C）］等都可以满足大规模蛋白质组分析的需求。其中我们所使用的 Protean II xi（20cm×20cm）的运行条件分两个步骤，第一步每块胶限流 20 mA 运行 40 min，此时 IPG 胶上的蛋白质应进入二维胶，第二步为每块胶限流 30 mA，运行至溴酚蓝前沿到达二维胶

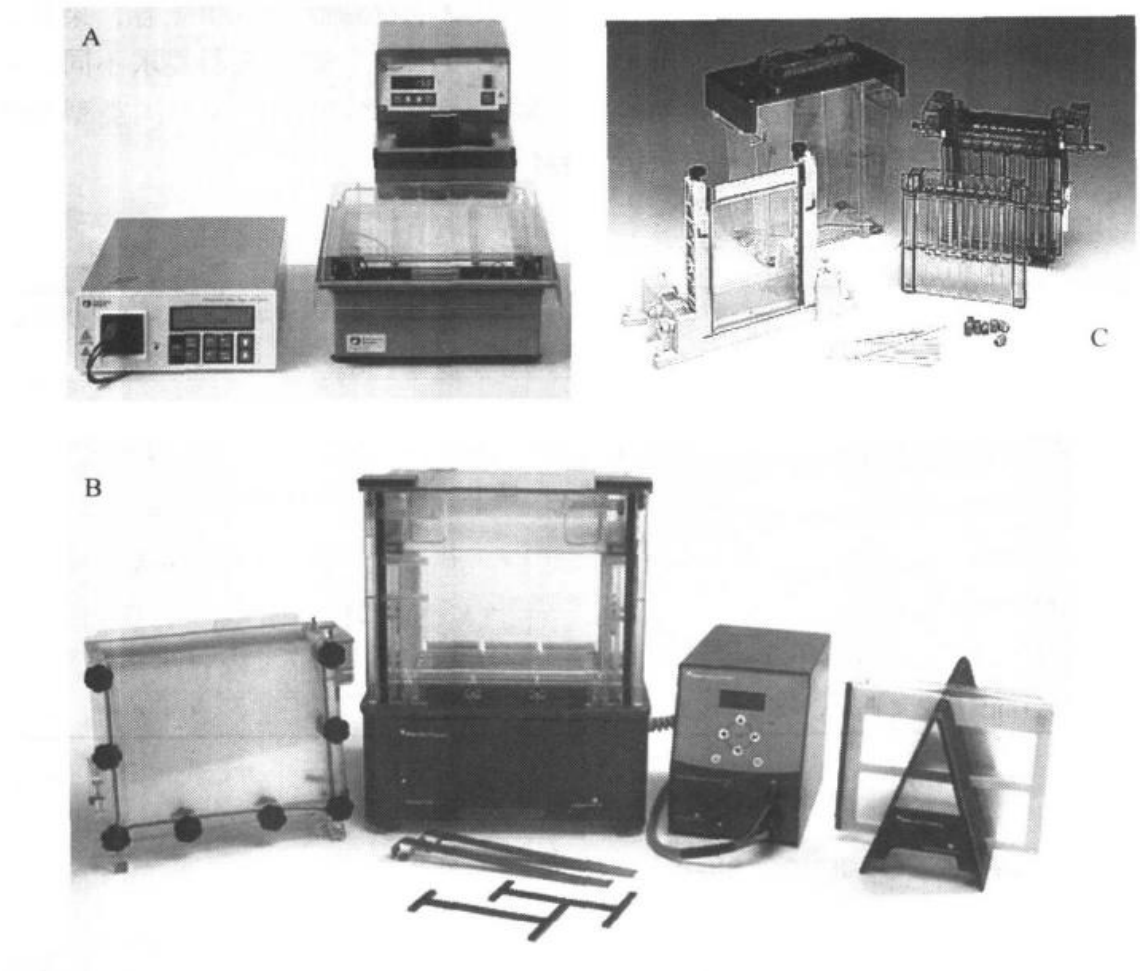


图 4-7　几种常用电泳仪及其附件

A. APB 公司水平电泳仪 Multiphor Ⅱ及其附件。从左至右分别为 EPS 3501 XL 电源，Multiphor Ⅱ水平电泳单元，后面为 Multi Temp Ⅲ恒温循环仪。B. APB 公司垂直电泳仪 Ettan Dalt Ⅱ及其附件。从左至右依次为灌胶模具、电泳槽、电源、玻璃支撑架。C. Bio-Rad 公司垂直电泳仪 Protean Ⅱ xi Cell 电泳槽、灌胶模具及附件。

底部大约需要 5 h，下一步的工作是凝胶的显色。

第四节　胶上蛋白的检测

用于二维电泳分离蛋白质常规检测和定量的普遍方法是考马斯亮蓝染色和银染[20]，但测序技术和质谱灵敏度的提高，尤其是最新的 Q-TOF 类液相色谱串联质谱技术的发展，大大提高了质谱鉴定的灵敏度，从而使蛋白质显色的地位从简单的蛋白质点呈现转换为集成化的蛋白质微量化学表征过程中的关键步骤。随着高灵敏度蛋白质分析方法和电泳后鉴定技术的结合，考马斯亮蓝染色和银染用于蛋白质组研究的局限性日益暴露出来。新的染色技术如荧光标记、同位素标记技

术在提高了灵敏度的同时，也与自动化的蛋白质组平台切胶技术相兼容，染色技术向高灵敏度和自动化方向发展。由于不同染色方法对设备和仪器要求不同，在实验中要综合考虑各种因素，以选择最佳方案。表 4-4 给出了各种不同类型染色方法的灵敏度、对软硬件的要求、质谱兼容性等性能指标。

表 4-4　不同染色方法的性能一览表

染色方法	灵敏度	活细胞应用	线性范围	与质谱兼容性	试剂耗费	对软硬件需求（扫描系统及其他）
考马斯亮蓝染色	100 ng	no	3	+ +	+	+
负染	15 ng	no	3	+ -	+	+
银染	200 pg	no	7	+	+	+
荧光染色	400 pg	no	10^4	+ + +	+ + + +	+ + + +
荧光标记	250 pg	no	10^4	+ +	+ + +	+ + + +
磷触屏	0.2 pg	yes	10^5	+ + +	+ + +	+ + + +
稳定同位素标记	< 1 pg	yes	?	+ +	+ +	+ + + +

一、考马斯亮蓝（Commassie blue）染色

自从传统考染技术和甲醇/醋酸水溶液结合用于 PAGE 胶染色至今，至少已尝试了 600 种关于 PAGE 胶考染染色的线性和检测灵敏度的不同方法[21]。考马斯亮蓝染料可以检测到 30 ~ 100 ng 蛋白质，其灵敏度远低于银染或荧光检测[20]。但考染的优点是染色过程简单，所需配制的试剂少，操作简便，无毒性，染色后的背景及对比度良好，与下游的蛋白质鉴定方法兼容。在样品量允许的情况下，考染是一个比较好的选择。这儿列出一个根据文献修改后的考染程序[22]：

1）固定：50% 乙醇/10% 冰醋酸的水溶液，至少 30 min，可过夜；

2）染色：染色液为 45% 甲醇/10% 冰醋酸/0.25% 考马斯亮蓝 G250 溶液过滤所得，将凝胶浸泡后染色至少 2 h；

3）脱色：多次更换脱色液（25% 乙醇/8% 冰醋酸水溶液）直至背景脱净。为了加快脱色，可略加温度；

4）保存：脱色后凝胶经水洗后扫描备份，用保鲜膜包裹即可，通常在一个月内均可进行质谱鉴定，若需保存时间长，可先把需鉴定蛋白质点切下进一步脱色后于 −80 ℃冰箱保存。

PAGE 胶经 G250 胶体考马斯亮蓝染色后可以获得无背景或很低的背景[23,24]，估计其染色机制是因为在胶体染料和溶液中的自由扩散染料间形成平衡，溶液的少量自由染料穿透凝胶基质，有选择性地对蛋白质染色，而胶体态的

染料排阻在胶外，阻止了背景生成[23]。胶体考马斯亮蓝检测蛋白质的极限是8～10 ng[20]。染色过程通常是在高浓度的三氯醋酸、高氯酸或磷酸中进行，并辅以甲醇或乙醇。质谱对蛋白质修饰的研究表明，用含三氯醋酸和乙醇的考马斯亮蓝染色的蛋白质易导致谷氨酸羧基侧链的不可逆酸催化酯基化[25]，这样就使所得肽谱数据复杂化，但也可通过在分析软件中加入算法来考虑这种修饰。胶体考马斯亮蓝染色的缺点是所需时间较长，染色过程大约要24～48 h，这限制了该方法的应用。

二、银染

目前常用的适于质谱鉴定的凝胶蛋白染色方法为考马斯亮蓝染色，其灵敏度为100 ng，也就是说，对10 kDa的蛋白质，在考染胶上显色所需量为10 pmol，这显然与生物质谱的高灵敏度特性不相匹配，而且表达丰度低的蛋白质，用考染难以显色。银染是非放射性染色方法中灵敏度最高的，其灵敏度可达200 pg，由于银染成本较低，在目前仍然是差异蛋白质组分析中最常用的显色方法，但由于银染过程中醛类的特异反应，使得对凝胶酶切肽谱提取存在困难。大多数实验室的策略是采用银染凝胶进行图谱分析，然后加大上样量，进行考染并将凝胶切下用于下游鉴定。由于银染和考染的特异性并不相同，使得这两种不同的方法所得到的二维电泳图谱可比性不好，这样不利于蛋白质组研究的高通量筛选。改进银染方法，使之适于胶内酶切及质谱鉴定，已成为蛋白质组研究的迫切要求。

经典银染方法是将二维电泳凝胶在固定液中固定后，在含戊二醛的溶液中增敏，然后在$AgNO_3$溶液中浸泡，蛋白与Ag^+结合，凝胶空白背景中的Ag^+由于结合不牢，大部分被洗去，而含蛋白质的区域则由于蛋白质中自由氨基与Ag^+的相互作用，使这些位置的Ag^+没能被洗去，随后在碱性环境中，甲醛溶液将结合在蛋白带上的Ag^+还原为金属银，银颗粒沉积在蛋白点上，沉淀的银颗粒又产生自催化反应，提高银染反应灵敏度，使蛋白点显示棕黄色或棕黑色。已报道的有多于100种的银染方法，表4-5中列出了常见的几种方法，这些不同方法的存在也说明，没有一种方法是完美的。银染分酸性（$AgNO_3$染色）和碱性（氨银法）两种。酸性染色背景浅，容易控制，所以用的较多，方法也多种多样，而碱性染色则灵敏度稍高，但背景深，难以控制。在对二维电泳凝胶进行染色时，如果不加增敏剂戊二醛，胶内会出现大量的透明背景点，在该位置的蛋白质点不能被显色，估计其原因与一向胶条中所加入的载体两性电解质有关。戊二醛的作用是利用醛基的交联作用和反应活性，提高银染灵敏度，而交联作用和对银颗粒的固化作用又使肽段很难被洗脱出来。为了探索银染凝胶鉴定可能性，Shevchenko[27]等去掉戊二醛，在一向SDS-PAGE中使1.6 ng（25 fmol）BSA显色，并在不脱色的情况下实现对5ng BSA进行胶上酶切和MALDI-TOF-MS

表 4-5　$AgNO_3$ 染色方法列表

方法	1		2		3		4		5	
	试剂	时间	试剂	时间	试剂	时间	试剂	时间	试剂	时间
固定	乙醇 100 ml，乙酸 25 ml	30 min	甲醇 125 ml，醋酸 30 ml，37%甲醛 0.125 ml	≥1 h	甲醇 125 ml，醋酸 12.5 ml	20 min			甲醇 125 ml，醋酸 12.5 ml	
冲洗			50%乙醇	3×20 min	甲醇 125 ml	10 min			甲醇 125 ml	10 min
					蒸馏水	10 min			蒸馏水	10 min
增敏	乙醇 75 ml，戊二醛(50%) 0.625ml，$Na_2S_2O_3$(5%) 10 ml，NaAc 17g	30 min	$Na_2S_2O_3$(5%)1 ml	1 min	$Na_2S_2O_3$(5%)1 ml	1 min	乙醇 75 ml，戊二醛(50%) 0.625ml，$Na_2S_2O_3$(5%) 10 ml，NaAc 10.25 g	2 h	$Na_2S_2O_3$(5%) 2.5 ml，0.15%$K_3Fe(CN)_6$，0.75g Na_2CO_3	
冲洗	蒸馏水	3×5 min	蒸馏水	3×20 sec	蒸馏水	2×1 min	去离子水	2×3 min	去离子水	3×5 min
孵育	$AgNO_3$(2.5%) 25 ml，甲醛(37%) 0.1 ml	20 min	$AgNO_3$(2.5%) 20 ml，甲醛(37%) 0.188 ml		$AgNO_3$(2.5%) 10 ml	20 min 4℃	$AgNO_3$(2.5%) 10 ml，甲醛(37%) 0.025 ml	30 min	$AgNO_3$(2.5%) 10 ml	20 min
冲洗	蒸馏水	2×1 min	蒸馏水	2×20 sec	蒸馏水	2×1 min	去离子水	20s	去离子水	2 min
显色	Na_2CO_3 6.25 g，甲醛(37%) 0.05 ml	2～5 min	Na_2CO_3 15 g，37%甲醛 0.125 ml，$Na_2S_2O_3$(5%) 0.02 ml		Na_2CO_3 5 g，37%甲醛 0.1 ml		Na_2CO_3 6.25 g，37%甲醛 0.025 ml，pH10.9	20 min	Na_2CO_3 5 g，37%甲醛 0.05 ml	
冲洗			蒸馏水	2×2min						
终止	EDTA-Na_2·$2H_2O$ (3.65g)	10 min	甲醇 125 ml，醋酸 30 ml	10 min	5% 醋酸		EDTA(50mmol/L)		5%醋酸	
冲洗	蒸馏水	3×5 min	甲醇 125 ml	≥20 min	1% 醋酸		EDTA (50mmol/L) 0.02% thimerosal	Dark, 4℃		
贮存						4℃		Dark, 4℃		
参考文献	A-P Silver Staining kit		Blum H[26]		Shevchenko A[27]		Jungblut PR[28]	Dark, 4℃	in situ Digestion Protocol from Kratos	

PMF 鉴定，达到 36% 的覆盖率，只是质谱图的噪声太大。Gharahdaghi[29]等对上文的实验方法加以改进，采用硫代硫酸钠（$Na_2S_2O_3$）和铁氰化钾[$K_3Fe(CN)_6$]脱色。$K_3Fe(CN)_6$ 在碱性环境中具有氧化性，可在 $Na_2S_2O_3$ 溶液中将金属银氧化为银离子并形成亚铁氰化银，$Na_2S_2O_3$ 可以与亚铁氰化银结合形成水溶性复合物，从而去除蛋白质点上的银粒子。将二者的贮备液在临用前混合，银染的蛋白条带浸泡其中，振荡混匀 2～3 min，可以看到胶条的棕黄色逐渐褪去。他们的实验结果证明，脱色后的质谱图优于不经过脱色而得到的图谱。而 Yan[30]则去除增敏液中的戊二醛和孵育液中的甲醛，进行探索。可以看到，目前银染方法改进的焦点集中在对银染过程中醛类的取舍问题。我们对 50 ng 兔磷酸化酶 b 标准蛋白在有和无戊二醛的条件下银染条带的结果进行胶内酶切和肽质量指纹谱鉴定，结果也证实了去除戊二醛后质谱分析能获得较好的图谱，数据库检索也获得了明确的结果（图 4-8）。目前多数文献中的蛋白质点的鉴定已开始采用银染胶内酶切质谱鉴定，方法基本建立在 Shevchenko 和 Gharahdaghi 等的研究基础上。

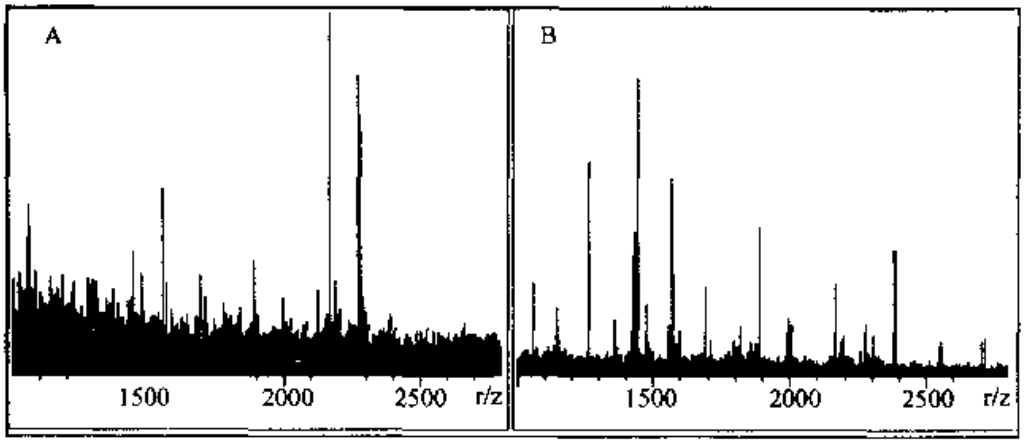


图 4-8　兔磷酸化酶 b 银染凝胶肽质量指纹图谱

A. 酸性染色含戊二醛。B. 酸性染色不含戊二醛。上样量均为 50 ng。

三、负染

标准考染方法所存在的一个缺点是染色步骤中含蛋白质固定过程，银染法除了固定过程外，还有增敏步骤，这些步骤会将蛋白质提取至胶外，从而减少蛋白质产量，使得用于随后微量化学表征的样品量更少了。负染的开发则是专门为了提高 PAGE 胶上蛋白质的回收率。常用的负染方法有铜染[31]、锌-咪唑[32,33]等，其灵敏度稍高于考染，可达到 50ng 以下。负染结果在凝胶表面会产生半透明背景，其中蛋白质以透明的形式被检测出来。凝胶显示黑色背景或相应的底色[34]。

染色过程很快，约需 5～15 min 即可完成，蛋白质的生物学活性可以得到保持。一旦用络合剂如 EDTA 或 Tris/甘氨酸转移缓冲液络合金属离子，就可进行提取来转移蛋白质[35]。负染适用于蛋白质显色、整个蛋白质的胶上被动提取以及质谱分析[36]。

四、荧光染色

由于考染、负染技术灵敏度不够，而银染的线性很差，在蛋白质组研究特别是比较蛋白质组学研究中的应用有限。荧光试剂显色对蛋白质无固定作用，与质谱兼容性好，而其灵敏度与银染相仿，但线性范围要远高于银染，由于上述原因，使二维电泳分离蛋白质的荧光检测正受到普遍关注和应用，特别是在致力于大规模蛋白质组研究的实验室。几种新的荧光染料也显示出了与集成化蛋白质组学平台相结合的良好前景。

最近比较新的应用是利用丙基 Cy3 和甲基 Cy5 两种染料分别对两个不同蛋白质样品进行荧光标记，并在一块 2-D 胶上同步运行，由于两种修饰后染料的激发波长不同，这样在同一块胶上可以用两个波长范围进行扫描，得到的两个图像经软件匹配即可很方便地找到两个样品的差异[37]。由于是在同一块胶上运行，完全避免实验因素对重复性的影响。上述方法的缺点包括需要对蛋白质进行共价修饰、改变了标记蛋白质的移动性能和由于荧光探针光猝灭作用所引起的快速衰减。蛋白与蛋白之间由于可被荧光团修饰的功能基团数目不同而使得灵敏度变化很大，荧光团的加入同时会降低蛋白质溶解性能[37]。可以通过标记样品中一小部分功能基团来补偿，但标记蛋白质与未标记蛋白质之间轻微的相对分子质量差异可能会导致分离后蛋白质的 Edman 降解和质谱鉴定出现偏差。此外，在运行 2-DE 或等电聚焦时，由前期荧光分子衍生可能导致异常蛋白质迁移[38]。

另外一种很受关注的荧光染料是 SYPRO Ruby，它是一种专利的基于钌的金属螯合染料。SYPRO Ruby Protein Blot 在狭缝印迹时其检测灵敏度为 0.25～1 ng/mm^2，常规大约 2～8 ng 的蛋白质可通过电转印检测，对照比较实验证实它具有和胶体金染色一样的灵敏度（AuroDye, APB, Arlington Heights, IL）。胶体金染色需要 2～4 h，而 SYPRO Ruby 染料染色在 15 min 内即可完成。SYPRO Ruby Protein Blot 染色的动态线性范围大大优于胶体金染色，扩展了约 1000 倍。该染料可用 300 nm 紫外透射仪激活或利用装配有 450 nm、473 nm、488 nm 乃至 532 nm 激光的成像系统。SYPRO Ruby 染料不干扰质谱或免疫检测过程，与 SYPRO Rose 不同，它对核酸不染色。

SYPRO Ruby 2-DE 和 IEF 染色仅需一步即可获得 PAGE 内蛋白质的低背景染色，而且不需要进行长时间脱色步骤。这些染料的线性动态范围扩展到超过三个数量级，在性能上超过了银染和考马斯亮蓝染色，对等电点从 3.5 到 9.3 范围

的 11 个蛋白质标准的评估表明，SYPRO Ruby IEF 凝胶染色的灵敏度比银染要高 3～30 倍[39]。银染结果很差的蛋白质往往用 SYPRO Ruby 染料很容易检测到。虽然比 SYPRO Orange 和 Red 染料更为灵敏，其最佳染色要稍慢一些，约需 4 h。SYPRO Ruby 染料与胶体考马斯亮蓝染料的性质类似，都是终点式（end-point）染色，因此，对染色时间的要求不严格，可以过夜，也不用担心会显色过度。虽然荧光染色是一种很有前景的染色方法，但由于 SYPRO Ruby 染料的代价很高，而且荧光凝胶所需扫描仪价格不菲，使得多数实验室还没有条件用荧光染色来取代银染、考染等方法，但随着各个国家在蛋白质组研究领域资金的大量投入，荧光染色及其相关设备必将进入更多的实验室。

第五节　存在问题

二维电泳虽然仍是蛋白质组研究不可替代的分离方法，但有些问题依然是该技术所无法解决的，主要存在于以下几个方面：① 低丰度蛋白质点的检测，有许多细胞因子及功能蛋白质往往表达量很低，用目前的显色方法难以呈现。而且按目前检测到的蛋白质点数，不可能全部包括细胞内所表达的上万个甚至更多蛋白质，那么，剩下的那部分蛋白如何检测；② 极酸和极碱性区蛋白的分离，按目前商业化的胶条，最多能分离 pH3～11 范围内的蛋白质，而且碱性区蛋白质的分离难度依然很大；③ 高分子质量区蛋白的分离，由于 IPG 胶条在重泡胀时，分子质量大于 100 000Da 的蛋白质很难进入胶条，这部分蛋白质也很难在二维电泳胶上呈现出来；④ 膜蛋白的提取和有效分离；⑤ 需要一种真正高通量的胶上蛋白鉴定技术。这些蛋白质的呈现问题成为蛋白质组研究的一个瓶颈，也使一些科学家对二维电泳的前景提出质疑。近年来的一些新技术也在致力于改善这种状况，如三步提取方法的探索大大提高了膜蛋白的提取效率（见第三章），一种被称为分子扫描仪的新型蛋白质鉴定技术也正在形成之中，该技术利用涂有酶解液的膜，在双向凝胶转膜的同时进行酶切，再利用质谱扫描鉴定，此方法有利于高通量鉴定技术的实现，相信在不久的将来就会投入应用[40,41]。Andrews 等则利用基质辅助的激光解吸飞行时间质谱仪取代第二向 SDS-PAGE 凝胶，他们直接将一向 IPG 胶与芥子酸饱和基质溶液共结晶后，用质谱测定一向胶条中每一个位置中所含蛋白质的相对分子质量，这样就得到一个含等电点、相对分子质量和信号强度（蛋白质丰度）三维坐标的图谱。通过比较显示，这种方法所得到的信息（蛋白质）要多于通过二维电泳分离并鉴定所得到的信息，通过这项新技术可以获得样品的“虚拟二维电泳谱（Virtual 2-D）”[42]。而其他一些人则在二维电泳以外的分离技术上力求有所突破，如一维 SDS-PAGE 分离-串联质谱技术[43]，二维色谱技术[44,45]及有针对性的免疫沉淀技术[43]等，其中最有希望的是二维色谱技术，因为利用这种技术可以避开相对分子质量和等电点的限制，目前的主要

问题是分辨率和重现性尚不能与二维电泳相比。因此，不管怎样，未来几年内二维电泳仍将在蛋白质组研究技术体系中扮演着重要角色，而且不排除还会在自动化和显色技术上取得更大的进展。

参考文献

1 郭尧君. 蛋白质电泳实验技术. 北京：科学出版社，1999.

2 何忠效，张树政. 生物化学实验技术丛书：电泳（第二版）. 北京：科学出版社，1999.

3 Hames B D, Richwood D. Gel Electrophoresis of Proteins (2th): The Practical Approach Series. Oxford. IRL Press, 1990. 4

4 Hjertén S. Arch Biochem Biophys Suppl, 1962, 1: 147

5 Smithies O, Poulik M D. Two-dimensional electrophoresis of serum proteins. Nature, 1956, 177: 1033

6 Dunn M J. Two dimensional polyacrylamide gel electrophoresis. In: Advances in Electrophoresis (ed. Chrambach, A., Dunn, M. J. and Radola, B. J.). Vol. 1. Weinheim. VCH, 1～109

7 O'Farrel P H. High resolution two-dimensional electrophoresis of proteins. J Biol Chem, 1975, 250 (10): 4007～4021

8 O'Farrel P Z, Goodman H M, O'Farrel P H. High resolution two-dimensional electrophoresis of basic as well as acidic proteins. Cell, 1977, 12 (4): 1133～1141

9 Bjellqvist B, Ek K, Righetti P G, Gianazza E, Gorg A, Westermeier R, Postel W. Isoelectric focusing immobilized pH gradients: principles, methodology, and some applications. J Biochem Biophys Meth, 1982, 6 (4): 317～339

10 Görg A, Postel W, Gunther S. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis, 1988, 9 (9): 531～546

11 Görg A, Obermaier C, Boguth G, Harder A, Scheibe B, Wildgruber R, Weiss W. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis, 2000, 21 (6): 1037～1053

12 Görg A, Boguth G, Obermaier C, Posch A, Weiss W. Two-dimensional polyacrylamide gel electrophoresis with immobilized pH gradients in the first dimension (IPG-Dalt): The state of the art and the controversy of vertical versus horizontal systems. Electrophoresis, 1995, 16 (7): 1079～1086

13 Bjellqvist B, Pasquali C, Ravier F, Sanchez J C, Hochstrasser D. A nonlinear wide-range immobilized pH gradient for two-dimensional electrophoresis and its definition in a relevant pH scale. Electrophoresis, 1993, 14 (12): 1357～1365

14 Görg A, Obermaier C, Boguth G, Csordas A, Diaz J J, Madjar J J. Very alkaline immobilized pH gradients for two-dimensional electrophoresis of ribosomal and nuclear proteins. Electrophoresis, 1997, 18 (3～4): 328～337

15 Görg A, Postel W, Gunther S, Weser J, Strahler J R, Hanash S M, Somerlot L, Kuick R. Approach to stationary two-dimensional pattern: influence of focusing time and immobiline/carrier ampholytes concentrations. Electrophoresis, 1988, 9 (1): 37～46

16 Görg A, Postel W, Friedrich C, Kuick R, Strahler J R, Hanash S M. Temperature-dependent spot positional variability in two-dimensional polypeptide patterns. Electrophoresis, 1991, 12 (9): 653～658

17 Herbert B. Advances in protein solubilization for two-dimensional electrophoresis. Electrophoresis, 1999, 20 (4～5): 660～663

18 Laemmli U K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature,

1970, 227 (259): 680～685

19 Dunn M J. Gel Electrophoresis: Proteins. Oxford: BIOS Scientific Publishers

20 Brush M. Dye hard: protein gel staining products. Scientist, 1998, 12: 16～22

21 Neuhoff V, Stamm R, Eibl H. Clear background and highly sensitive protein staining with Coomassie Blue dyes in polyacrylamide gels. Electrophoresis, 1985, 6: 427～448

22 郭尧君. SDS电泳技术的实验考虑及最新进展. 生物化学与生物物理进展, 1991, 18 (1): 32～37

23 Neuhoff V, Arold N, Taube D, Ehrhardt W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. Electrophoresis, 1988, 9 (6): 255～262

24 Neuhoff V, Stamm R, Pardowitz I, Arold N, Ehrhardt W, Taube D. Essential problems in quantification of proteins following colloidal staining with coomassie brilliant blue dyes in polyacrylamide gels, and their solution. Electrophoresis, 1990, 11 (2): 101～117

25 Haebel S, Albrecht T, Sparbier K, Walden P, Korner R, Steup M. Electrophoresis-related protein modification: alkylation of carboxy residues revealed by mass spectrometry. Electrophoresis, 1998, 19 (5): 679～686

26 Blum H, Beier H, Gross H J. Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. Electrophoresis, 1987, 8: 93～99

27 Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of protein from silver-stained polyacrylamide gels. Anal Chem, 1996, 68 (5): 850～858

28 Jungblut P R, Seifert R. Analysis by high-resolution two-dimensional electrophoresis of differentiation-dependent alterations in cytosolic protein pattern of HL-60 leukemic cells. J Biochem Biophys Methods, 1990, 21 (1): 47～58

29 Gharahdaghi F, Weinberg C R, Meagher D A, Imai B S, Mische S M. Mass spectrometric identification of proteins from silver stained polyacrylamide gel: A method for the removal of silver ions to enhance sensitivity. Electrophoresis, 1999, 20 (3): 601～605

30 Yan J X, Wait R, Berkelman T, Harry R A, Westbrook J A, Wheeler C H, Dunn M J. A modified silver staining protocol for visualization of protein compatible with matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry. Electrophoresis, 2000, 21 (17): 3666～3672

31 Lee C, Levin A, Branton D. Copper staining: a five-minute protein stain for sodium dodecyl sulfate-polyacrylamide gels. Anal Biochem, 1987, 166 (2): 308～312

32 Castellanos-Serra L, Proenza W, Huerta V, Moritz R L, Simpson R J. Proteome analysis of polyacrylamide gel-separated proteins visualized by reversible metative staining using imidazole-zinc salts. Electrophoresis, 1999, 20 (4～5): 732～737

33 Castellanos-Serra L, Hardy E. Detection of biomolecules in electrophoresis gels with salts of imidazole and zinc II: a decade of research. Electrophoresis, 2001, 22 (5): 864～873

34 Wirth P J, Romano A. Staining methods in gel electrophoresis, including the use of multiple detection methods. J Chromatogr A, 1995, 698 (1～2): 123～143

35 Castellanos-Serra L, Fernandez-Patron C, Hardy E, Huerta V. A procedure for protein elution from reverse-stained polyacrylamide gels applicable at the low picomole level: An alternative route to the preparation of low abundance proteins for microanalysis. Electrophoresis, 1996, 17 (10): 1564～1572

36 Cohen S L, Chait B T. Mass spectrometry of whole proteins eluted from sodium dodecyl sulfate-polyacrylamide gel electrophoresis gels. Anal Biochem, 1997, 247 (2): 257～267

37 Unlu M, Morgan M E, Minden J S. Difference gel electrophoresis: a single gel method for detecting changes

in protein extracts. Electrophoresis, 1997, 18 (11): 2071－2077

38 Urwin V E, Jackson P. Two-dimensional polyacrylamide gel electrophoresis of proteins labeled with the fluorophore monobromobimane prior to first-dimensional isoelectric focusing: imaging of the fluorescent protein spot patterns using a cooled charge-coupled device. Anal Biochem, 1993, 209 (1): 57－62

39 Steinberg T H, Chernokalskaya E, Berggren K, Lopez M F, Diwu Z, Haugland R P, Patton W F. Ultrasensitive fluorescence protein detection in isoelectric focusing gels using a ruthenium metal chelate stain. Electrophoresis, 2000, 21 (3): 486～496

40 Bienvenut W V, Sanchez J C, Karmime A, Rouge V, Rose K, Binz P A, Hochstrasser D F. Toward a clinical molecular scanner for proteome research: parallel protein chemical processing before and during western blot. Anal Chem, 1999, 71 (21): 4800－4807

41 Binz P A, Muller M, Walther D, Bienvenut W V, Gras R, Hoogland C, Bouchet G, Gasteiger E, Fabbretti R, Gay S, Palagi P, Wilkins M R, Rouge V, Tonella L, Paesano S, Rossellat G, Karmime A, Bairoch A, Sanchez J C, Appel R D, Hochstrasser D F. A molecular scanner to automate proteomic research and to display proteome images. Anal Chem, 1999, 71 (21): 4981－4988

42 Loo R R, Cavalcoli J D, VanBogelen R A, Mitchell C, Loo J A, Moldover B, Andrews P C. Virtual 2 D gel electrophoresis: visualization and analysis of the E. coli proteome by mass spectrometry. Anal Chem, 2001, 73 (17): 4063～4070

43 Heller M, Goodlett D R, Watts J D, Aebersold R. A comprehensive characterization of the T-cell antigen receptor complex composition by microcapillary liquid chromatography-tandem mass spectrometry. Electrophoresis, 2000, 21 (11): 2180－2195

44 Chong B E, Yan F, Lubman D M, Miller F R. Chromatofocusing nonporous reversed-phase high-performance liquid chromatography/electrospray ionization time-of-flight mass spectrometry of proteins from human breast cancer whole cell lysates: a novel two-dimensional liquid chromatography/mass spectrometry method. Rapid Commun Mass Spectrom, 2001, 15 (4): 291～296

45 Covert B A, Spencer J S, Orme I M, Belisle J T. The application of proteomics in defining the T cell antigens of Mycobacterium tuberculosis. Proteomics, 2001, 1 (4): 574～586

（应万涛）

第五章　图像分析与细胞蛋白谱的建立

第一节　蛋白电泳图像分析系统

获得蛋白质双向电泳凝胶后，要对凝胶图像进行备份保存，从而以数字化图像的形式存储下来，而且要尽量完整地保留定性和定量信息，以利于进一步的分析。通过对批量双向凝胶图谱的分析，应该获得的信息包括每一块凝胶中所分离得到的总蛋白质点数、双向电泳凝胶之间的批次重现性、蛋白质点的缺失和出现以及多块胶之间蛋白质点的表达丰度的定量变化，而这些工作仅仅依靠眼睛的分辨能力来执行是不现实的，很难系统全面地对整个凝胶图像进行判断，更不用说要对蛋白质点的表达量进行比较分析。要对上千个蛋白质点进行分析，只有依赖于双向电泳凝胶分析软件，完整的图像处理包括图像的采集和双向电泳图像分析(图 5-1)。

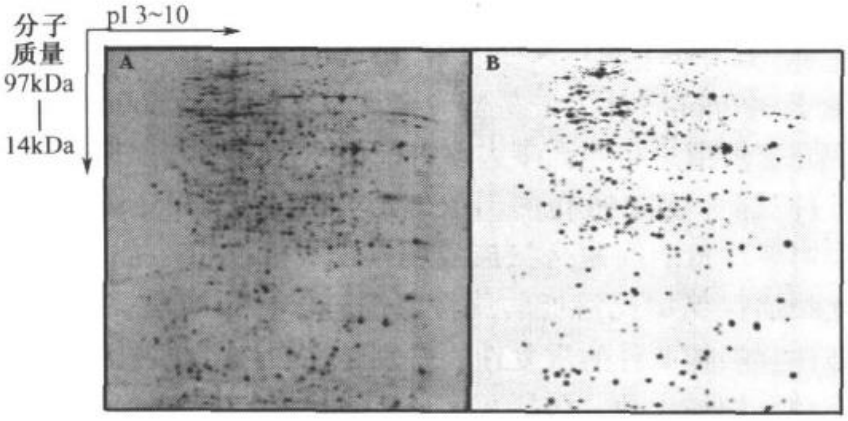


图 5-1　图像处理前后凝胶对比

A. 人肺巨细胞双向电泳图谱，垂直 SDS-PAGE（14%连续胶）。

B. 经 PDQuest 软件背景消减和高斯拟合处理后的图谱。

一、二维凝胶电泳图像采集

常用的图像采集系统有电感偶合装置（charge coupled devices，CCD）、光密度扫描仪、激光诱导荧光检测器等。无论何种采集系统，都必须具备透射扫描的功能，因为与反射方式不同，透射扫描时灵敏度较高，可以根据吸光度的大小获

得蛋白质点的光密度信息，一般来说，该光密度值与蛋白质点的表达丰度成正比，因而保持了图像的真实性，使图像中包含了更丰富的信息，以便于软件分析时的定量比较。由于下一步的图像分析基本上依赖于各个蛋白质点的光密度值相对大小，因此，图像采集质量的好坏关系到图像分析结果的可靠性。影响图像采集质量的因素包括扫描系统的分辨率、灵敏度，以及扫描时所选择的图像对比度和明亮度。

目前常用的可见光扫描仪如 Amersham Pharmacia Biotech 公司 ImageScanner、Bio-Rad 公司的 GS-710 光密度仪等，可以扫描考马斯亮蓝染色、银染和负染凝胶，扫描方式有透射和反射两种，可以选择黑白、256 色和真彩等模式。荧光、化学发光和磷屏扫描仪如 Amersham Pharmacia Biotech 公司 Typhoon 系列扫描仪，可以选择在可见光范围以外的其他激发波长扫描凝胶，这些凝胶通过肉眼无法观察到颜色，但由于通过呈现方法所得到的蛋白质点在鉴定时无需脱色处理，与下游质谱鉴定的兼容性较好，适用于蛋白质组的自动化和高通量化，使 Typhoon 类的扫描仪越来越被看好，缺点是该类仪器的费用高，而且荧光染色的试剂费用也很高。上述几类扫描仪的分辨率均可达到 1000dpi（dot per inch）以上，但设置 dpi 值太大，图像保真度虽然提高了，而图像所占字节数也大大增加，使软件在分析图像时对处理器的要求提高，这样就大大延长了运行时间。在实际应用时，仪器分辨率设置为 300dpi、8bits 深度为宜，此时一块 20cm×20cm 的双向电泳胶（Protean II，Bio-Rad）保存为 tif 格式的文件后，约占用 6Mb 的字节空间，分辨率低于 300dpi 时会失去某些光密度信息，高于 300dpi 时，分析软件所能识别的图像光密度信息得不到多大改善，而图形文件大大加大。由于目前所用的凝胶分析软件只能识别 256 色的图像，要求在扫描时设定图像色彩为 256 色。

图像采集时要注意的问题是扫描强度校正（intensity calibration），定期的强度校正可以使扫描仪保持光密度的线性范围，保证扫描系统所获得图像的真实性。同时亦要注意扫描条件的重复性，特别是在图像分析时需要进行相互匹配和定量比较的凝胶，扫描参数，包括灰度值、对比度、亮度设定一定要一致。

二、二维电泳凝胶图像分析

如果双向电泳实验的目的是要鉴定某一蛋白质混合物中所呈现出的全部蛋白质点，通过软件分析所要作到的有以下几点，包括蛋白质点数的统计、蛋白质点的定位、编号和相对丰度分析。如果是要进行差异蛋白质组分析，则需要对相互对照样品的凝胶图像进行同步分析，比较对应蛋白质点的表达丰度，获得差异蛋白质点的缺失、出现以及表达量的变化等信息，进而获得有意义的蛋白质点，所有这些工作都需要通过图像分析软件来完成。双向电泳凝胶分析软件的工作原理是围绕每一个象素点构建一个算子（operator），然后比较算子中心和边缘的光密

度值，比值达到一定强度时该中心点的象素即被认为是蛋白质点的一部分。重复此过程，软件即可自动检测出完整的蛋白质点。经过多年的发展，目前已形成了比较完善的图像分析系统。图像分析软件有 melanie II & 3[1,2]、PDQuest 6.0、ImageMaster 2D Elite 3.10、Phoretix 2D[3]等，它们的基本功能是相近的，都包括以下几个步骤[4]（图 5-2）：

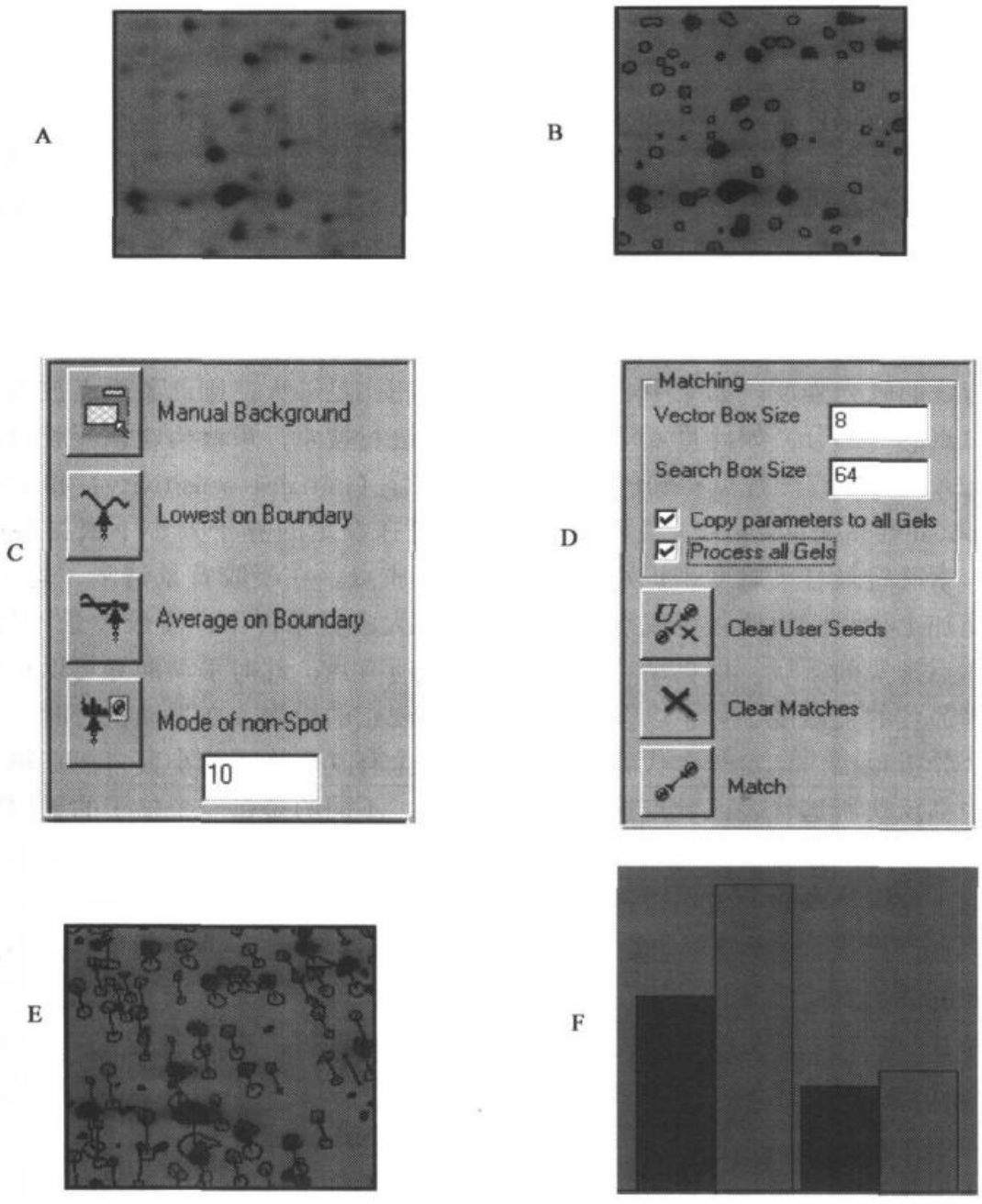


图 5-2　二维电泳图像分析流程

A. 扫描后的原始图像。B. 蛋白质点的检测。C. 几种不同的背景消减方法。D. 凝胶匹配参数的选择。E. 匹配后图像。F. 归一化处理后的相对点体积。

1）蛋白质点检测（spot detection）；

2）背景消减（background subtraction）；

3）归一化处理（normalization）；

4）蛋白质点匹配（spot matching）。

下面以 ImageMaster 2D Elite 3.10 分析软件为模板来解释一下图像分析的基本原理，其思路基本上适用于目前大多数双向电泳图像分析软件。

1．蛋白质点检测

点的检测包括胶内点的定位、点的形状确定及点的体积（丰度）和面积的计算，该步骤是以后其他各步分析的基础，如果在后续分析中发现点检测不完善，重新执行点检测步骤的话，所有后续分析的信息将被清除，所以必须认真对待点的检测，尽量能检测全面，因此它也是很耗费时间的一步。为了减短处理时间，检测前预先要用鼠标圈定要处理点所在的区域，处理区域越小，后续分析所耗费的时间就越短。点检测原理是以每一个象素点为中心构建算子，然后比较算子中心与边缘的吸光强度，如果中心与边缘的强度相比足够大，则此象素点被认为是某蛋白点的一部分，如此重复对每一个象素点进行搜索，最终构建该凝胶的图谱（图 5-2A，B）。有两种点检测的方法，自动执行（initialise sensitivity）和手动执行（spot detection wizard）。自动执行时软件将在设定的算子大小下根据凝胶背景等因素确定最佳灵敏度来识别凝胶中的蛋白质点。在凝胶背景清晰、均匀时，此功能加上少量编辑即可完成点的检测。但在大多数条件下，凝胶背景不均匀，特别在点密集的区域，用自动功能很难检测到所有点，此时手动向导可以让分析者选择最佳的灵敏度、算子以及噪声参数，实现点的检测。在手动执行时，首先在整个分析范围内选定一个代表区域，进入一个有 9 个格子的小窗口，在每一个格子中都反映所选代表区域点检测的执行情况，不同的是每一个格子的灵敏度和算子大小不同，选择最满意的格子，或直接调节下侧和右侧的灵敏度和算子大小，缩小这两个参数的大小范围，直到中间格子的检测结果达到最佳而其他的格子效果近似为止。进入下一步，就是消除背景，可以调节要消除背景的大小，最后得到的就是点检测结果。手动检测也带来了一个问题，就是如何保证多块胶之间的参数一致性。如果参数相差较大，检测到的点的数目也就相差较大，这样所获得的分析结果就失去了双向电泳的重复性。正确的参数选择需要经验的积累，同时也需要充分利用点检测中的手工编辑功能（如下所述）。

无论是自动检测还是手动检测，所得到的结果都不可能一步到位，可能仍有一些点未被识别出来，还有一些污染物等造成的“假点”，另外一些因距离过近被识别成一个点，或因邻近点的强度太高而被识别为背景掩盖，对这些不足之处需要进行进一步的编辑。ImageMaster 软件提供了画点、删除点、提高峰值、边缘增加、点切割等进一步的编辑功能（图 5-3），以对软件所直接识别出来的信

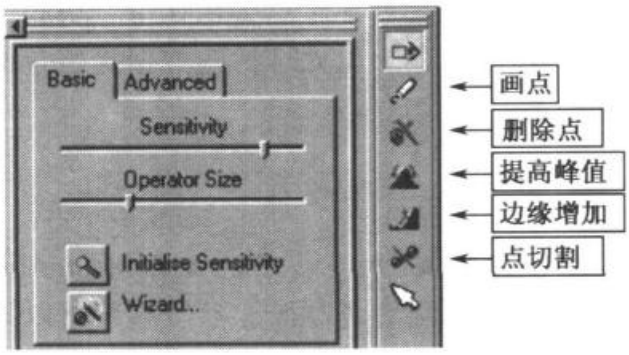


图 5-3　点检测工具（ImageMaster 2D Elite）

息进行补充。

2．背景消减

点检测完成后，所得到的蛋白质点的体积（volume）是一个绝对值，由构成该蛋白质点的各个象素吸光强度的值累加得到，在凝胶背景较深时（实际上大部分银染凝胶和考染凝胶都很难控制使之无背景），这些背景值将会被叠加到点体积中去，因而使蛋白质点体积与实际值相比要略微偏高。一些图像分析软件如PDQuest 可在分析前对图像作平滑、对比度调节和背景消减等处理。ImageMaster 软件在点检测后进行背景消减，共有 4 种模式可供选择：手工消减（人工选取认为可以充当背景的蛋白质点或区域，软件将以这些点的光密度值均值作为凝胶背景）、边界最低强度消减（沿被检测的蛋白质点边界取强度最低的点为背景点）、边界平均象素值消减（取被检测点边界点的平均强度为背景）、非点模式消减（图 5-2C），后 3 种均是完全自动的，其中非点模式消减首先通过软件找到完全包含该完整蛋白质点的长方形，然后将该长方形按设定的象素数扩大，该象素在图 5-2C 中被设定为 10，软件会自动检查该区域内不包含在任一蛋白质点内的所有象素点，出现频率最高的象素强度将被认为是该蛋白质点的背景。该功能在选择蛋白质点周边象素点强度为背景的同时，消除了将毗邻的蛋白质点识别为背景的可能性，是经常选用的背景消减方式。如果上述 4 种方式均不能满足要求，也可以直接编辑某个点的背景水平。

3．归一化处理

该步骤对点的体积进行归一化处理，以便于对不同胶上的同一蛋白质点进行准确客观的相对定量。归一化处理一般有 2 种模式，一种模式基于全部点体积之和，以每一个点的体积除以总和，通过这种模式得到的数值可能会很小，因此可以乘以一个放大因子，当该放大因子的值为 100 时，所得到的数值即为点体积百分含量。另外一种归一化的模式是基于某一点，该点可能是一已知的标志蛋白或

人为添加的数值，此时要注意的是该标志点必须在每一块胶上都存在，否则将无法在不同胶之间进行对比匹配。归一化处理后的数据是否合理决定着图像分析数据统计学分析的可靠性。

4．蛋白质点匹配

匹配是2D图像分析中很重要的一步，在ImageMaster软件中，这也是一个交互式的过程。匹配时首先要创建参考凝胶，参考凝胶可以是要分析比较的一组凝胶中的一张，也可以是几张凝胶合并而成的平均胶。用户可以通过改变向量框（vector box）和搜索框（search box）的大小来操纵匹配。由于电泳过程中的一些影响因素的作用，即使是同一样品两次电泳图像之间也存在位置的偏差，在这种情况下，可以调整上述两个参数，得到较好的匹配效果（图5-2C，D，F）。如果凝胶变形太大，自动匹配功能无法找到匹配的点，可以进行人工匹配。选择明显对应的蛋白质点，通过鼠标的点击可以使两个对应点匹配，该蛋白质点被识别为种子点（user seeds）。一旦选择种子点后，可选择自动匹配功能，软件将自动按照种子点匹配矢量的大小和方向对种子点附近的蛋白质点进行匹配，通过这种方法可以实现变形较大的凝胶之间的匹配。如果一个种子点不行，还可以添加多个，前提是一定要保证种子点的对应关系正确无误。

蛋白质点匹配以后，软件可以自动对相互之间匹配上的蛋白质点进行比较分析，可以设定差异表达的范围，如表达量提高两倍，软件会自动列出表达量提高两倍的蛋白质点，通过这种方法，可以获取差异表达的信息。同时，也可以让软件标示出蛋白质点有和无的变化。

5．其他处理

除以上基本功能外，各软件还包括等电点及相对分子质量校正、数据表格的输出以及与internet网络数据库的接口实现等功能。配套软件ImageMaster 2D databasc可以进行统计处理和进化树分析，PDQuest软件本身则具有回归分析和统计分析等功能。胶上蛋白质点通过等电点和相对分子质量的校正后所获得的信息，可以为下一步的网上蛋白质搜索鉴定提供重要的参考因素，从而排除其他范围蛋白质所造成的干扰。此外，这些图像分析软件很重要的功能是能直接和Internet上的数据库连接，有利于网上数据的检索以及电泳图谱的对比分析。ImageMaster 2D Elite 3.10还支持网页制作功能，方便了网上数据库的构建。

6．结果输出

双向电泳图像分析会生成大量的数据，这些数据可以以表格的形式输出或粘贴到Microsoft Excel软件中去，利用该软件的分析功能进行分析。检测结果还可生成凝胶报告或蛋白质点报告，包括图像和列表，并可存储为tif文件。利用

Excel 软件可以对每一个蛋白质点在胶上的纵横坐标的变化进行统计分析，从而得到凝胶之间位置的差异，对双向电泳的重复性作出评价[5]。通过对匹配结果进行统计分析，可以查出没有匹配（unmatched）的蛋白质点的编码，排除掉实验干扰因素后，一部分没有匹配的点就可能是在实验组中新出现的蛋白质。

7．存在问题

双向电泳图像分析软件虽然在智能性、背景消减、光密度校正等方面都尽量优化以客观地呈现出实验结果，但仍然存在不足之处。首先是处理时间，要完整全面地检测到一块胶上的所有蛋白质点，大约需要 1 h 的时间，如果是处理一批凝胶，大量的时间要花费在点的检测和匹配上，特别是匹配阶段。有些软件虽然可以设定自动化批处理功能，整个流程让软件自动执行，但由于目前双向电泳凝胶很难达到完全重现，使得这些功能也只能是停留在理论水平，不可能在实际中应用。图像分析软件未来的发展方向，应该是在双向电泳技术进一步完善的基础上，提高其定量识别、自动化处理和智能分析等能力。

第二节　二维电泳蛋白谱数据库

双向电泳 IPG 技术的应用，提高了电泳图谱的重复性，使得实验室之间的合作与图谱对比成为可能，Internet 网络的普及及上网实验室的逐渐增多，给人们的合作及数据共享提供了方便，微量测序、肽质量指纹谱、氨基酸组成分析和串联质谱技术的发展，加快了蛋白鉴定的步伐。在此基础上，大量的蛋白质组双向电泳表达谱得以构建[6-12]，为了使这些数据能得到迅速有效的管理、共享和利用，一些研究单位就专门构建了双向电泳蛋白质组数据库。由于不同的数据库构建者理念不同，导致网上电泳数据库的使用方法各异，针对这种混乱性，Appel 等于 1996 提出了一个联合数据库的构建标准，概括了网上联合数据库的 5 个规范[6]：

1）通过远程关键词搜索可以调出目的蛋白质；

2）数据库可以通过超文本交叉参考而与其他数据库链接；

3）除了单个可查寻数据库外，从每一个蛋白链接点，需要链接一个不同交叉参考数据库的目录；

4）通过点击胶上蛋白点可以调出单个蛋白相关信息；

5）双向电泳分析软件可以设定直接进入联合电泳数据库。

这 5 条规范目前已成为默认的准则。除了利用网页制作工具来编写数据库外，也可以从网上下载双向电泳数据库制作软件包 Make2ddb package[7,8]来创建标准的联合数据库。

目前网上已公布了大量的蛋白质组双向电泳数据库，仅瑞士生物信息研究所

的 WORLD-2DPAGE 网站上就链接了包括不同种类和种属的双向电泳数据库网站共 39 个[6]，其中 SWISS 2-D PAGE 数据库就包括了人、鼠的器官、细胞及酵母、大肠等不同生物来源的 30 多类双向电泳数据库，表 5-1 列出了 WORLD-2DPAGE 中所包含的一些重要网站内容及网址。国内研究单位也紧跟发展，如上海生化所筹划建立的肝癌和钩端螺旋体蛋白质组数据库，军事医学科学院基因组与蛋白质组中心筹划建立的胎肝和肺癌蛋白质组数据库，中国医学科学院肿瘤所筹划建立的食管癌蛋白质组数据库等。

表 5-1　WORLD-2DPAGE 中链接的部分双向电泳数据库网站内容及网址

数据库名称	种　属	网站中所包含的组织或细胞	网　址
SWISS-2D PAGE	Human	Liver, Plasma, HepG2, Red blood cells, Lymphoma, HepG2 Secreted Proteins, Cerebrospinal Fluid, Macrophage Like Cell Line, Erythroleukemia Cell, Platelet, Kidney, Promyelocytic Leukemia Cells, Colorectal epithelia cells, Colorectal adenocarcinoma cell line (DL-1), Soluble nuclear proteins and matrix from liver tissue, Nucleoli from HeLa cells (1D and 2D)	http://www.expasy.ch/ch2d/
	Mus musculus	Liver, Gastrocnemius muscle, Pancreatic islet cells, Brown adipose tissue, White adipose tissue, Soluble nuclear proteins and matrix from liver tissue	
	Arabidopsis thaliana, *Dictyostelium discoideum*, *Escherichia coli* (*several pI ranges*), *Saccharomyces cerevisiae*		
SIENA-2 DPAGE	Human	Breast ductal carcinoma and histologically normal tissue, Amniotic fluid	http://www.bio-mol.unisi.it/2d/2d.html
	Chlamydia	*trachomatis*, *Caenorhabditis elegans*	
2-DE database	Human	Jurkat T-cells	http://www.mpiib-berlin.mpg.de/2D-PAGE/
	Mycobacterium bovis, *Mycobacterium tuberculosis*, *Helicobacter pylori*, *Borrelia garinii*		
Human and mouse 2-D PAGE databases	Human	Keratinocytes, Transitional cell carcinomas, Bladder squamous cell carcinomas, Urine, Fibroblasts	http://biobase.dk/cgi-bin/celis
	Mouse	Kidney	
HSC-2DPAGE	Human	Heart (ventricule), endothelial cells	http://www.hare-field.nthames.nhs.uk/nhli/protein
	Rat, Dog	Heart (ventricule)	
PDD Protein Disease Database	Human	Plasma, Cerebrospinal Fluid, Urine	http://www-lecb.ncifcrf.gov/PDD/

续表

数据库名称	种　属	网站中所包含的组织或细胞	网　址
PPDB PhosphoProtein DataBase	Murine	Lymphocytes	http://www-lecb.ncifcrf.gov/phosphoDB/
Washington University Inner Ear Protein Database	Human	Perilymph	http://oto.wustl.edu/thc/innerear2d.htm
	Guinea pig	Endolymph, organ of Corti, otoconia, tectorial membrane	
Toothprint	Rat	Enamel	http://biocl11.otago.ac.nz:591/tooth/home.htm
Rat Serum Protein	Study Rat Group	Serum (control Sprague-Dawley (CD) female and male rats, turpentine-treated rats, diabetic rats, stroke-prone rats), urine (control female and male rats, stroke-prone rats), cerebrospinal fluid (control and stroke-prone rats)	http://linux.farma.unimi.it/
Rat Serum Protein Database	Rat	Serum (normal and acute phase)	http://depts.washington.edu/~ruedilab/info/ratserum/ratserum.html
Mito-pick	Human	Mitochondria	http://www-dsv.cea.fr/thema/MitoPick/Mito2D.html
Human Colon Carcinoma Protein Database	Human	Placenta, Colorectal cell line	http://www.ludwig.edu.au/jpsl/jpslhome.html
SSI-2DPAGE	*Mycobacterium tuberculosis*		http://www.ssi.dk/en/forskning/tbimmun/tbhjemme.htm
Parasite proteome maps	*Caenorhabditis elegans*, *Fasciola hepatica*, *Hymenolepis diminuta*, *Monezia expansa*, *Necator americanus*		http://www.aber.ac.uk/~mpgwww/Proteome/ProtMaps.html
COMPLUYEAST-2DPAGE	*Candida albicans*, *Saccharomyces cerevisiae*		http://babbage.cnb.uam.es/2d/
YPD Yeast Protein Database			http://www.proteome.com/databases/
ANU-2DPAGE	Rice anther and Medicago truncatula		http://semele.anu.edu.au/2d/2d.html

参考文献

1 Appel R D, Palagi P M, Walther D, Vargas J R, Sanchez J C, Ravier F, Pasquali C, Hochstrasser D F. Melanie II—a third-generation software package for analysis of two-dimensional electrophoresis image: I. Features and user interface. Electrophoresis, 1997, 18 (15): 2724～2734

2 Appel R D, Vargas J R, Palagi P M, Walther D, Hochstrasser D F. Melanie II—a third-generation software package for analysis of two-dimensional electrophoresis images: II. Algorithms. Electrophoresis, 1997, 18 (15): 2735～2748

3 Mahon P, Dupree P. Quantitative and reproducible two-dimensional gel analysis using Phoretix 2D Full. Electrophoresis, 2001, 22 (10): 2075～2085

4 贾宇峰，林秋霞，郭尧君，郭鹤，刘少君. 蛋白质双向电泳图像分析. 生物化学与生物物理进展，2001，28 (2)：246～250

5 Corbett J M, Dunn M J, Posch A, Gorg A. Positional reproducibility of protein spots in two-dimensional polyacrylamide gel electrophoresis using immobilised pH gradient isoelectric focusing in the first dimension: an interlaboratory comparison. Electrophoresis, 1994, 15 (8 9): 1205～1211

6 Oh J M, Brichory F, Puravs E, Kuick R, Wood C, Rouillard J M, Tra J, Kardia S, Beer D, Hanash S. A database of protein expression in lung cancer. Proteomics, 2001, 1 (10): 1303～1319

7 Fountoulakis M, Juranville J F, Berndt P, Langen H, Suter L. Two-dimensional database of mouse liver proteins. An update. Electrophoresis, 2001, 22 (9): 1747～1763

8 Evans G, Wheeler C H, Corbett J M, Dunn M J. Construction of HSC 2DPAGE: a two-dimensional gel electrophoresis database of heart proteins. Electrophoresis, 1997, 18 (3 4): 471～479

9 Ji H, Reid G E, Moritz R L, Eddes J S, Burgess A W, Simpson R J. A two-dimensional gel database of human colon carcinoma proteins. Electrophoresis, 1997, 18 (3-4): 605～613

10 Ostergaard M, Rasmussen H H, Nielsen H V, Vorum H, Orntoft T F, Wolf H, Celis J E. Proteome profiling of bladder squamous cell carcinomas: identification of markers that define their degree of differentiation. Cancer Res, 1997, 57 (18): 4111～4117

11 Seow T K, Ong S E, Liang R C, Ren E C, Chan L, Ou K, Chung M C. Two-dimensional electrophoresis map of the human hepatocellular carcinoma cell line, HCC-M, and identification of the separated proteins by mass spectrometry. Electrophoresis, 2000, 21 (9): 1787～1813

12 Ou K, Seow T K, Liang R C, Ong S E, Chung M C. Proteome analysis of a human heptocellular carcinoma cell line, HCC-M: an update. Electrophoresis, 2001, 22 (13): 2804～2811

13 Appel R D, Bairoch A, Sanchez J C, Vargas J R, Golaz O, Pasquali C, Hochstrasser D F. Federated two-dimensional electrophoresis database: a simple means of publishing two-dimensional electrophoresis data. Electrophoresis, 1996, 17 (3): 540～546

14 http://www.expasy.org/ch2d/2d index.html

15 Hoogland C, Baujard V, Sanchez J C, Hochstrasser D F, Appel R D. Make2ddb: a simple package to set up a two-dimensional electrophoresis database for the World Wide Web. Electrophoresis, 1997, 18 (15): 2755～2758

16 http://us.expasy.org/ch2d/make2ddb.html

（应万涛）

第六章　生物质谱技术与蛋白质鉴定

质谱分析法是通过测定样品离子的质荷比（m/z）来进行成分和结构分析的分析方法。自 1886 年 Goldstein 发明早期质谱仪器常用的离子源，到 1942 年第一台单聚焦质谱仪商品化，质谱基本上处于理论发展阶段。随后质谱在电离技术和分析技术上的发展和完善，使之很快应用于地质、空间研究、环境化学、有机化学、制药等多个领域。然而，即使在等离子体解吸（plasma desorption，PD）和快原子轰击（fast atom bombardment，FAB）两项软电离质谱技术出现以后，质谱分析的相对分子质量也只是在几千道尔顿左右。真正意义上的变革以 80 年代中期出现的两种新的电离技术：电喷雾电离和基质辅助激光解吸电离为代表，这两种技术所具有的高灵敏度和高质量检测范围，使得在 fmol（10^{-15}）乃至 amol（10^{-18}）水平检测相对分子质量高达几十万道尔顿的生物大分子成为可能，从而开拓了质谱学一个崭新的领域——生物质谱，促使质谱技术在生命科学领域获得广泛的应用和发展[1]。

随着人类基因组全序列的问世，以大规模分析细胞或生物体内的蛋白质为主要内容的蛋白质组研究成为后基因组时代的重要研究内容。蛋白质组研究所必需的高灵敏度、高鉴定速度使得传统的生物化学方法远远不能满足要求，而生物质谱技术的发展，为蛋白质组研究提供了关键的技术手段，成为蛋白质组研究的三大关键技术之一[2,3]。

第一节　基质辅助激光解吸电离飞行时间质谱

一、MALDI 离子源

MALDI（matrix-assisted laser desorption ionization）的电离方式于 1988 年由 Karas 和 Hillenkamp[4] 提出。MALDI 的基本原理是将分析物分散在基质分子（尼古丁酸及其同系物）中并形成晶体，当用激光（337nm 的氮激光）照射晶体时，基质分子吸收激光能量，样品解吸附，基质-样品之间发生电荷转移使样品分子电离。基质的使用是这一电离技术的关键，基质吸收了激光的大部分能量并气化，同时将样品分子带入气相，样品分子只吸收少量激光能量，避免了分子化学键的断裂。基质在样品离子形成过程中充当了质子化或去质子化试剂，使样品分子带上正电荷或负电荷。

常用的基质是一些小分子有机酸及其衍生物，能够很好地吸收激光能量（见表 6-1）。

表 6-1 常用基质

基质简称	中文名称	英文名称	激发波长
SA	芥子酸 (3,5－二甲氧基－4－羟基肉桂酸)	sinapinic acid (3,5－dimethoxy－4－hydroxycinnamic acid)	337nm
DHB	龙胆酸 (2,5－二羟基苯甲酸)	gentisic acid (2,5－dihydroxybenzoic acid)	355nm
CCA	α－氰基－4－羟基肉桂酸	α－cyano－4－hydroxycinnamic acid	
PA	吡啶甲酸	picolinic acid	
3HPA	3－羟基吡啶甲酸	3－hydroxypicolinic acid	

MALDI 源使用的激光是脉冲式，每一脉冲激光产生的一批离子得到一张质谱图，一般使用的质谱图是多次脉冲激光扫描质谱峰结果的累加。MALDI 源产生的离子多为单电荷离子，质谱图中的谱峰与样品各组分的质量数有一一对应关系，因此 MALDI-MS 最适合分析多肽及蛋白质混合物。MALDI 源的离子化效率非常高，所以这种仪器是现今灵敏度最高的质谱仪，能够对极微量样品（fmol-amol）进行分析。

二、线性飞行时间质量检测器

MALDI-MS 通常用飞行时间（time of flight，TOF）检测器作为质量分析器，所构成的仪器称为基质辅助激光解吸/电离飞行时间质谱（MALDI-TOF-MS）。飞行时间检测器的质量分析范围很大，从几百道尔顿到 100 kDa 都可以检测。

离子源中每一脉冲激光产生的离子先经过一个加速电场，获得动能，再进入一个高真空无电场飞行管道，离子在此无电场飞行管道内以在加速电场获得的速度飞行。质量较轻的离子飞行速度快，较早到达检测器，较重的离子飞行速度慢，较晚到达检测器，离子的飞行时间与其质荷比的平方根 $(m/z)^{1/2}$ 成正比，通过检测飞行时间来测定离子的质荷比（图 6-1）。

由 $E = U \cdot z = \frac{1}{2} m v^2$ 和 $t = \frac{L}{v}$

推出 $t = const \cdot \sqrt{\frac{m}{z}}$

其中，E 为离子在加速电场获得的动能，U 为加速电场电压，z 为离子所带电荷，m 为离子质量，v 为离子飞行速度，L 为飞行管道长度，t 为离子飞行

时间，*const* 为一常数。

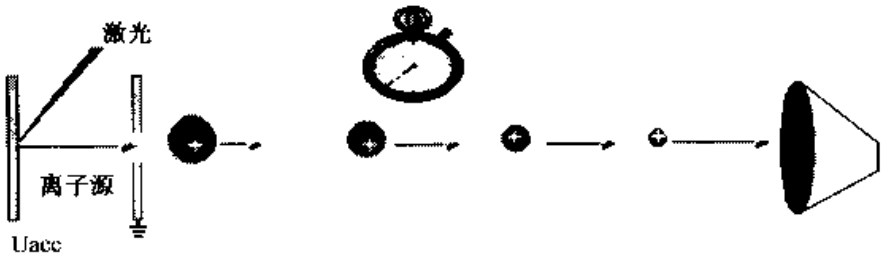


图 6-1　线性 MALDI-TOF 原理图

三、反射飞行时间质量检测器

线性飞行管的 MALDI-TOF-MS 的质量分辨率和测量准确度不高，测量误差在 2‰左右，不能满足蛋白质鉴定的需求。为了改进仪器的分辨率和质量测量准确度，在仪器的飞行管道内加了一个反射电场，称为 ion mirror 或 reflector，构成的仪器称为反射飞行时间质谱（RETOF-MS）（图 6-2、6-3）。反射电场起了能量聚焦作用，即具有相同质荷比但能量有细微差异的离子在反射电场作用下飞行时间达到一致，同时离子改变了飞行方向，延长了飞行距离，所以反射飞行时间质谱（RETOF-MS）的分辨率和质量测量准确度都有了很大提高。为了使离子能够转向并飞到反射检测器，反射电场的电压要大于加速电场电压。反射检测方式只能分析质量数 10 kDa 以下的离子。

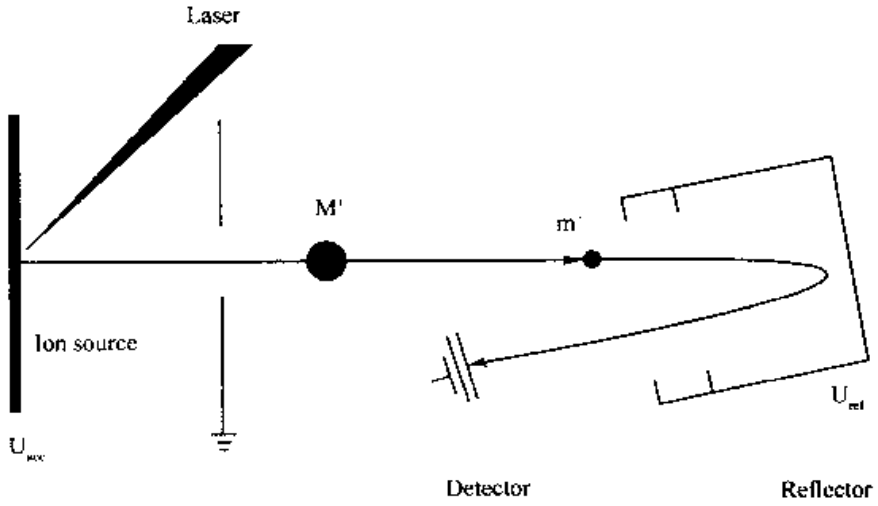


图 6-2　反射 MALDI-TOF 原理图

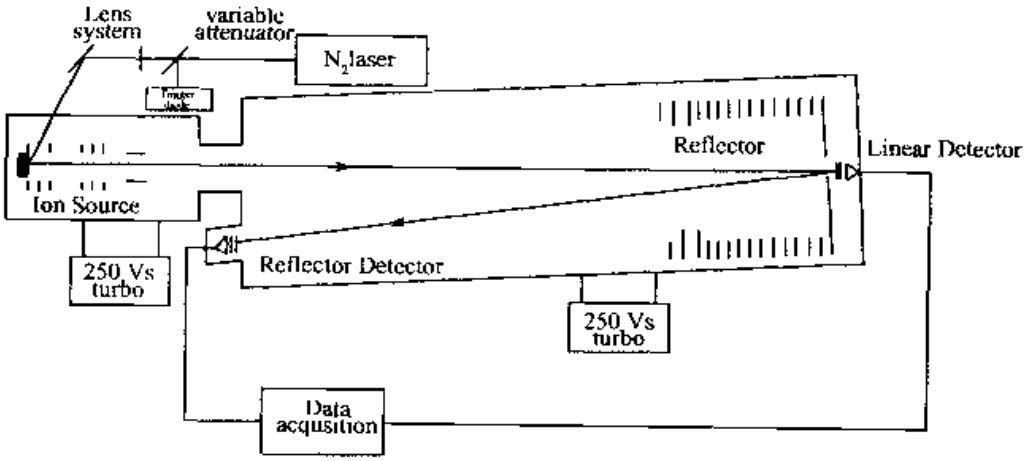


图 6-3　反射式 MALDI-TOF-MS 仪器结构图

具有反射飞行时间分析器的 MALDI-TOF-MS 可以分析源后衰变（post-source decay，PSD）产生的亚稳离子，这是一种串联质谱分析，可获得离子的结构信息。源后衰变即 MALDI 源产生的离子（母离子）在飞行管道飞行过程中发生裂解，丢失部分中性碎片，剩余的离子片段（子离子）成为质量减少、飞行速度不变，即动能减少的亚稳离子。母离子和子离子在反射电场中的通过时间不同，由于二者速度相同，质量较小，即动能较小的子离子通过时间较短，这个时间与离子质量成正比，与反射电场的电压成反比。在正常分析母离子时，子离子不能进入反射检测器。为了分析子离子，在实际操作中，通过逐步减小反射电场电压的方式检测子离子。PSD 谱图通常是由 10 到 20 个不同反射电压下的谱图叠加而成，为了覆盖质量在母离子质量的 10%～100% 范围内的子离子，反射电场电压要下降到原反射电压的 10%。

四、样品制备方法

要获得一张满意的质谱图，样品的制备十分关键。对蛋白质和多肽类样品，通常采用的样品制备方法如下：

基质 α-氰基-4-羟基肉桂酸用 0.1% 三氟乙酸、30%～50% 乙腈溶液配成饱和溶液，稀释 1 到 3 倍可以减少基质加和峰，要求每天新鲜配制。

样品溶于 0.1% 三氟乙酸溶液。

基质与样品以 1∶1 体积混合，取 0.5～1 μl 加到不锈钢样品靶上，在空气中放置几分钟，待溶剂挥发干后将样品靶送入仪器。

图 6-4 是由 MALDI-MS 分析得到的乙醇脱氢酶的胰蛋白酶酶切肽图。

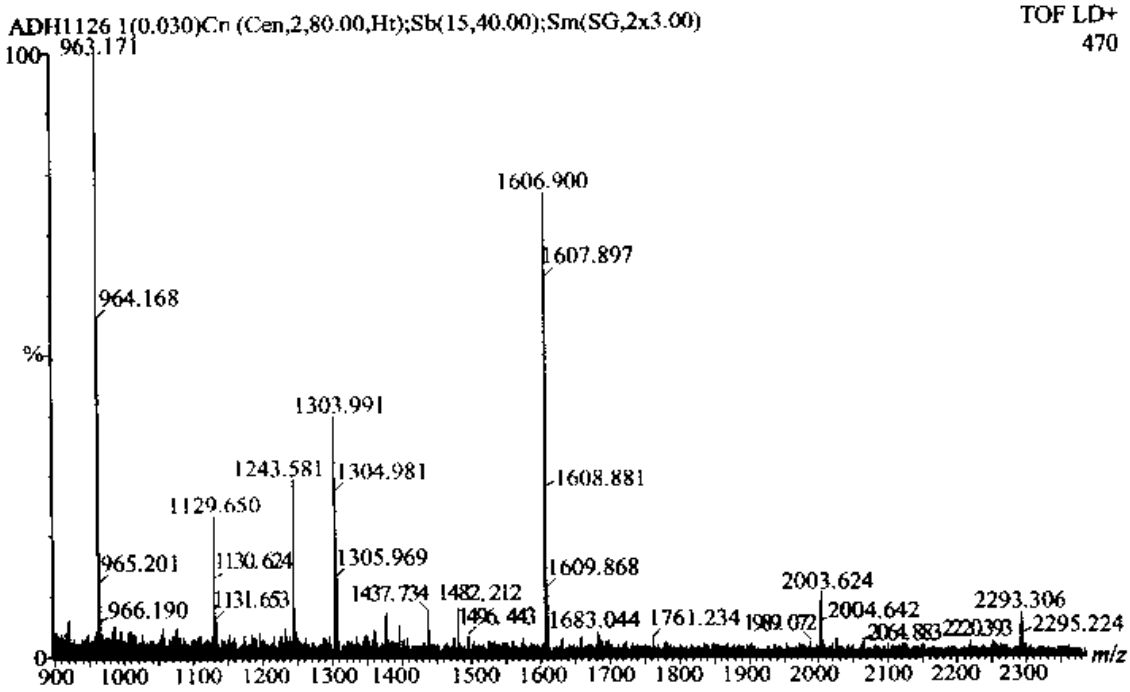


图 6-4　MALDI-MS 分析乙醇脱氢酶的胰蛋白酶酶切肽图

第二节　液相色谱-电喷雾-串联质谱

一、电喷雾电离源

电喷雾电离方法[5,6]是利用高电场使质谱进样端的毛细管柱流出的液滴带电，在 N_2 气流的作用下，液滴溶剂蒸发，表面积缩小，表面电荷密度不断增

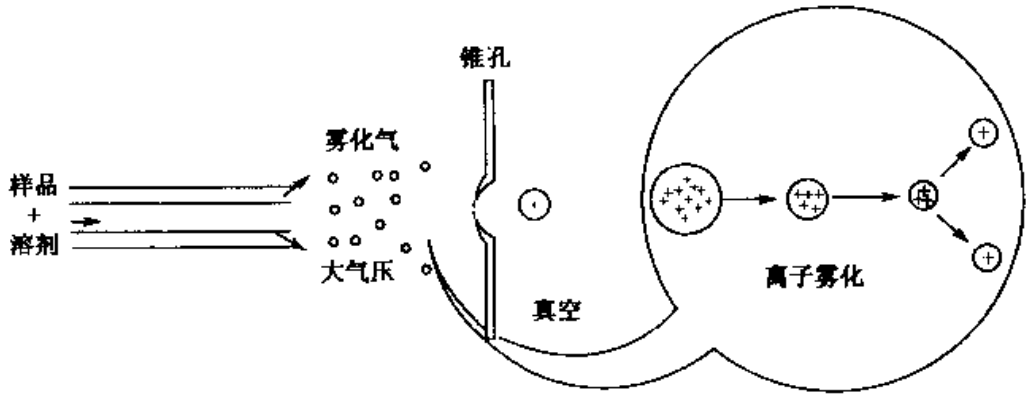


图 6-5　电喷雾电离原理图

加，直至产生的库仑斥力与液滴表面张力达到雷利极限，液滴爆裂为带电的子液滴，这一过程不断重复使最终的液滴非常细小呈喷雾状，这时液滴表面的电场非常强大，使分析物离子化并以带单电荷或多电荷的离子形式进入质量分析器（图6-5）。

电喷雾产生的生物大分子离子，如多肽、蛋白质等，经常带电荷数10个以上，形成 $[M+nH]^{n+}$ 系列多电荷离子（n 为质子化数，即电荷数），质荷比大大减小，弥补了四极杆等质量分析器测量范围窄的缺点。电喷雾电离产生的质谱图显示的是离子带不同电荷数的一系列质荷比峰，要根据这些峰值换算成离子质量。

已知 $m_j=(m+n_j)/n_j$

$m_k=(m+n_k)/n_k$

$n_j=n_k+1$（质子化，即加 H^+ 导致的离子化）

其中 m_j 为带电荷数为 n_j 的离子的质荷比，m_k 为带电荷数为 n_k 的离子的质荷比，m 为离子质量，可以推出

$$n_j=(m_k-1)/(m_k-m_j) \quad 及 \quad n_k=(m_j-1)/(m_k-m_j)$$

$$m=m_j\cdot n_j-n_j \quad 或 \quad m=m_k\cdot n_k-n_k$$

所以，根据相差一个电荷数的相邻两个多电荷离子峰的质荷比，可计算出多电荷离子所带电荷数，进而换算出该离子带单电荷时的质量。由图6-6A中的多电荷离子质荷比计算电荷数，峰659.8所带电荷数为

$$(714.7-1)/(714.7-659.8)=13$$

峰714.7所带电荷数为

$$(659.8-1)/(714.7-659.8)=12$$

换算成单电荷分子离子质量为

$$714.7\times12-12=8564.4$$

$$659.8\times13-13=8564.4$$

一般根据多个多电荷峰计算出多个质量值后取平均值。

分辨率高的质谱可以分辨出单个离子峰的同位素峰。根据相邻同位素峰的质荷比之差，也可以算出此峰代表的离子所带电荷数，即 $n=1/\triangle m$，其中 n 为该离子所带电荷数，$\triangle m$ 为相邻两同位素峰的质荷比之差。如图6-6B所示，相邻两同位素峰的质荷比之差为0.0769，则该峰所带电荷数为1/0.0769=13。根据电荷数可以计算出该多电荷离子的中性分子质量，即 $m=n\times m_{avg}-n$，其中 m_{avg} 为多电荷离子的平均质荷比。图6-6A中的多电荷离子峰转换后的分子质量为 $13\times659.8-13=8564.4$Da。

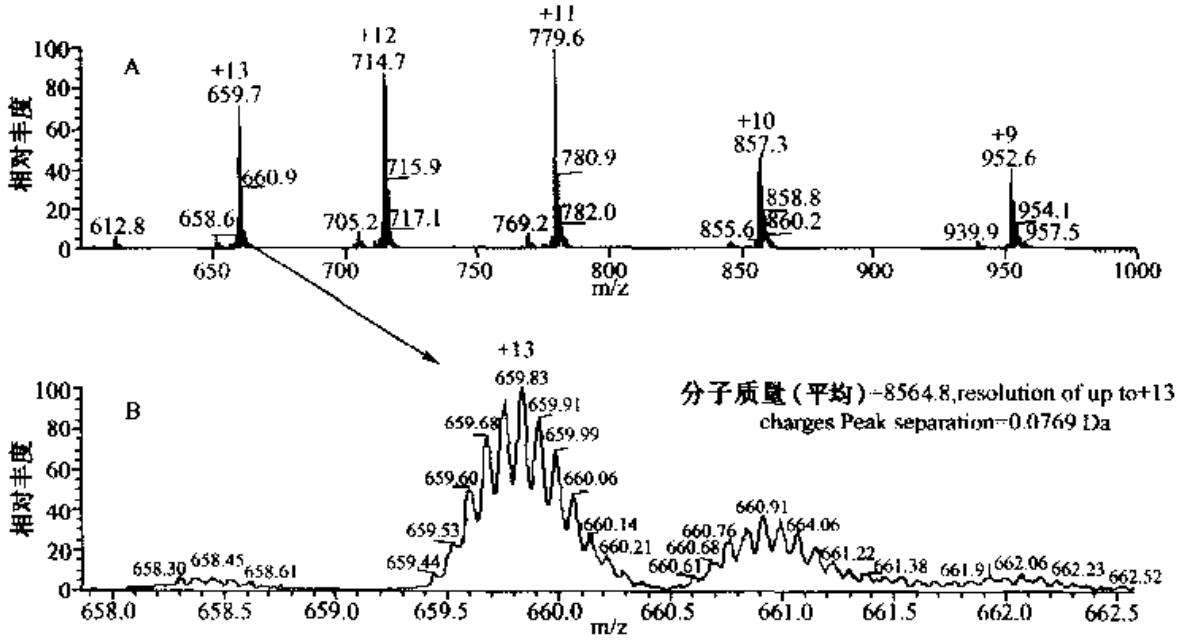


图 6-6　电离产生的多电荷离子峰及同位素峰

二、液相色谱-电喷雾-串联质谱（LC-ESI-MSIMS）

电喷雾离子源后连接串联质谱可检测离子结构碎片的质荷比，能够提供离子的结构信息。应用于多肽、核酸的测序及蛋白质的鉴定和翻译后修饰分析。

串联质谱仪[7]主要由离子源、多级质量分析器及碰撞室组成。第一级质量分析器起质量过滤器的作用，即从总离子谱中挑选出需进一步进行结构分析的母离子（precursor or parent ion）进入碰撞室，母离子在碰撞室内经高流速惰性气体碰撞诱导离解（collision induced desorption，CID）产生碎片离子/子离子（product or daughter ion），由第二级质量分析器分析子离子的质荷比。通过母离子和子离子的质量，研究二者关系及母离子的裂解规律，可以获得母离子的结构信息。

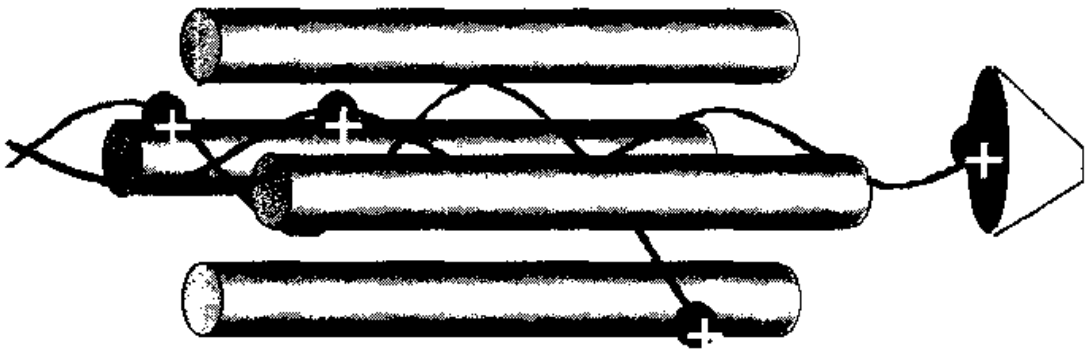


图 6-7　四极杆质量分析器

生物质谱仪常用的多级质量分析器有三级四极杆分析器（triple-quadrupole analyzer）、离子阱分析器（ion trap analyzer）和四极杆-飞行时间质量分析器（quadrupole-time of flight analyzer）。

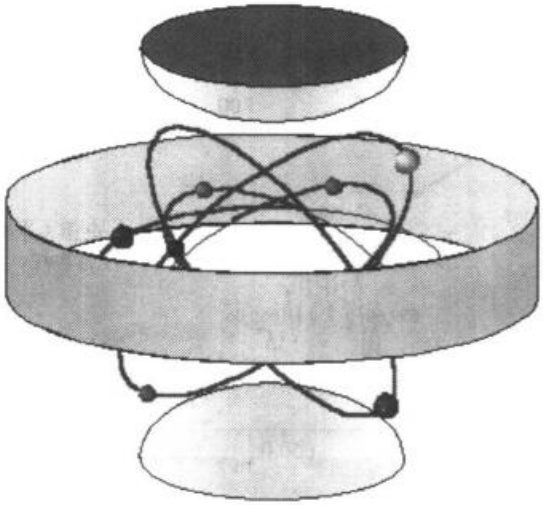


图 6-8　离子阱质量分析器

三级四极质量分析器由三个顺序连接的四极杆质量分析器（图 6-7）组成，第一级和第三级四极杆分析器均可用于离子选择（母离子扫描和子离子扫描）和

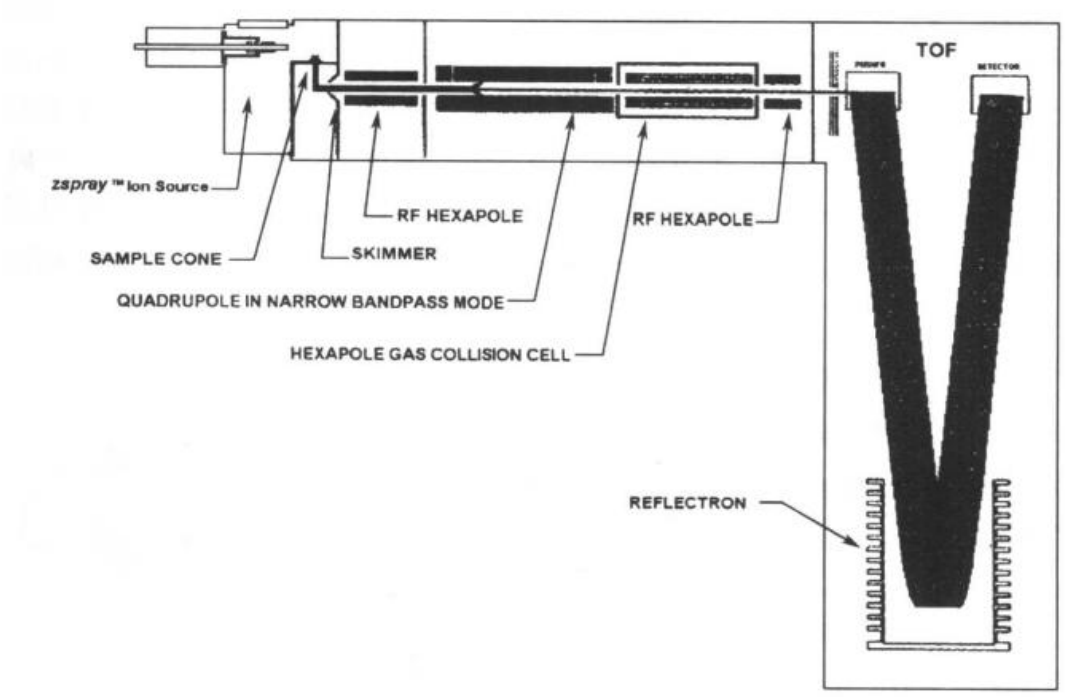


图 6-9　ESI-Q-TOF 仪器结构图

质量分析，中间的四极杆用于碰撞解离产生子离子。

离子阱分析器由三个双曲面电极（一个环形，两个斗笠形）围成一个离子捕集室（图 6-8），在环形电极射频（RF）电压的作用下，离子聚集在捕集室中心运动，在斗笠形电极上加电压可以选择出不同共振频率的离子进入检测器，也可以选择只有一种离子留在捕集室，再用高速惰性气体碰撞诱导解离产生子离子，进行串联分析。所以，离子阱分析器集一级、二级分析器及碰撞室于一体，甚至可进行（MS)^n 级串联分析。

四极杆-飞行时间分析器是由四极杆分析器和飞行时间分析器串联构成，四极杆分析器用于离子选择，其后有一个六极杆射频场用于离子的碰撞，诱导解离产生子离子，母离子和子离子的检测均由最后的飞行时间检测器完成。四极杆-飞行时间质量分析器和电喷雾离子源相连构成的仪器称 ESI-Q-TOF（图 6-9）。

电喷雾离子源可以方便地与高效液相色谱（HPLC）技术在线联用，在分析复杂混合物时很有优势，可以先分离再分析。液相色谱-电喷雾-串联质谱分析肽混合物鉴定蛋白质时，可以对每一个肽段进行序列分析，综合所有 MS 和 MS/MS 数据鉴定蛋白质，大大提高了鉴定准确度。图 6-10 是用 LC-ESI-MS/MS 分析肽混合物的实例，图 6-10A 是肽混合物的总离子流色谱图，图 6-10B 是肽混合物的总离子流质谱图，图 6-10C 和 6-10D 是其中两个肽段的串联质谱图。

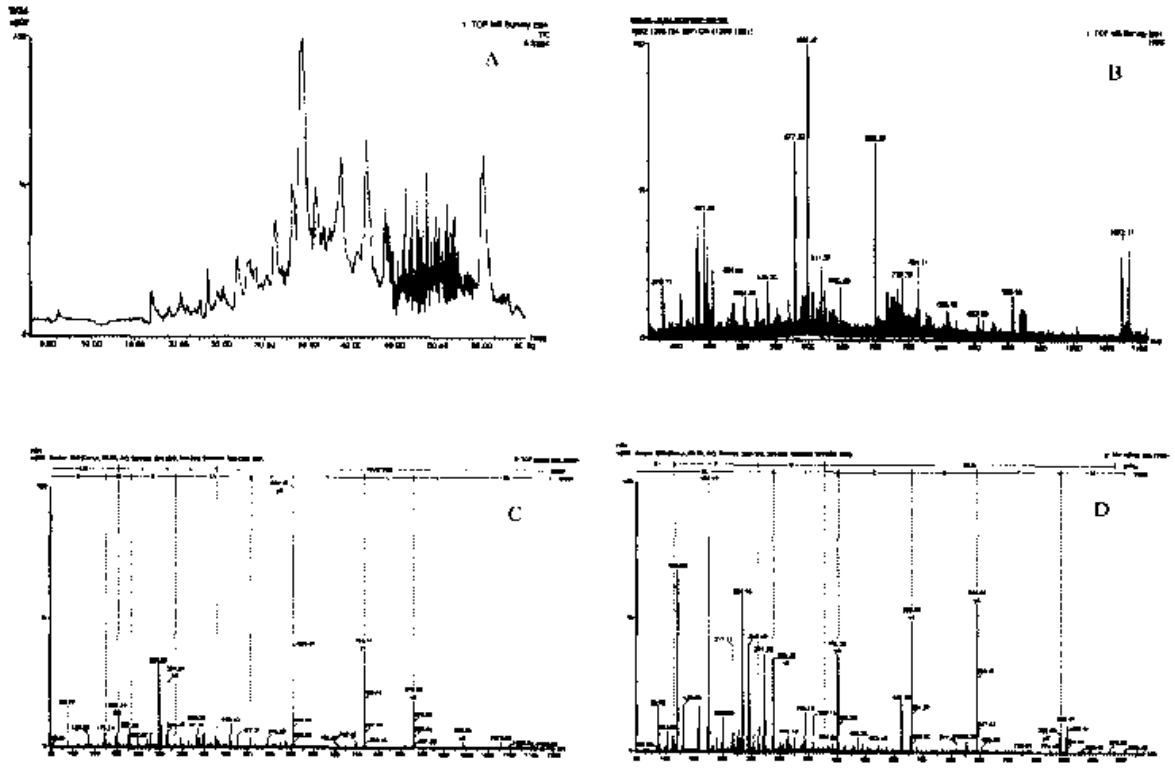


图 6-10　LC-ESI-MS/MS 分析肽混合物

A. 肽混合物的总离子流色谱图。B. 肽混合物的总离子流质谱图。C、D. 其中两个肽段的串联质谱图。

第三节　肽质量指纹谱鉴定蛋白质技术

一、肽质量指纹谱鉴定蛋白质原理

蛋白质的鉴定方法[8]主要是利用蛋白质的各种属性参数（attribute parameter）如相对分子质量、等电点、序列、氨基酸组成、肽质量指纹谱等在蛋白质数据库中检索，寻找与这些参数相符的蛋白质。如果在数据库中找不到，有可能是发现了新蛋白，要进行序列分析，合成 DNA 探针来表达、分离和鉴定这一蛋白。利用生物质谱技术可获得多种蛋白质属性参数，在蛋白质组鉴定中起着非常重要的作用[9,10]。

用肽质量指纹谱（peptide mass fingerprinting，PMF）鉴定蛋白质是目前蛋白质组研究中较为常用的鉴定方法，这一技术于 1993 年被多个研究小组分别独立提出[11,12]。肽质量指纹谱是指蛋白质被酶切位点专一的蛋白酶水解后得到的肽片段质量图谱。由于每种蛋白质的氨基酸序列（一级结构）都不同，蛋白质被酶水解后，产生的肽片段序列也各不相同，其肽混合物质量数亦具特征性，所以称为指纹谱，可用于蛋白质的鉴定，即用实验测得的蛋白质酶解肽段质量数在蛋白质数据库中检索，寻找具有相似肽指纹谱的蛋白质。电泳胶上的蛋白质可被原位酶切或转印到 PVDF 膜上酶切得到肽混合物，经质谱分析得到 PMF，检索数据库进行鉴定。还可以选取肽混合物中的某一肽段通过串联质谱测序鉴定，这部分内容将在下一节中予以介绍。基本路线如图 6-11 所示。

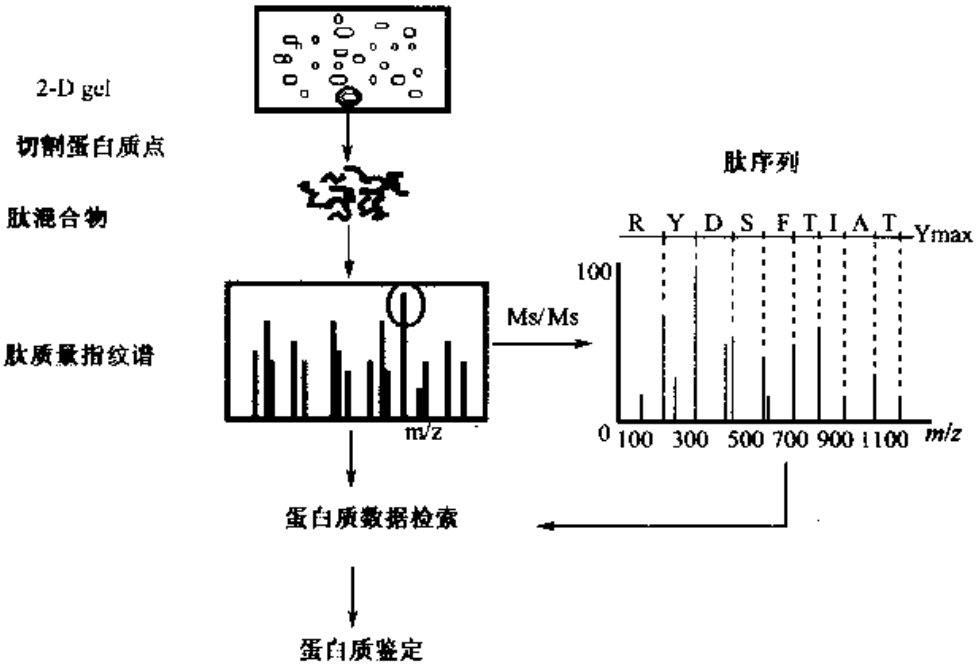


图 6-11　PMF 鉴定蛋白质基本路线

测定肽混合物质量数最有效的质谱仪是 MALDI-TOF-MS，灵敏度高，谱峰简单，每个谱峰代表一种肽段。但是蛋白质可能存在翻译后修饰，电泳过程中某些氨基酸会引入质量修饰（如半胱氨酸烷基化，甲硫氨酸氧化），这样在蛋白质的 PMF 中就会有一些质量数与理论值不符，但 PMF 鉴定的好处就在于不需要将全部肽质量数都与理论值相符。PMF 方法可同时处理许多样品，是大规模鉴定的首选方法。

用于制备肽谱的酶需具以下特点：酶的水解位点专一；酶切产生的蛋白的肽段大小适合质谱分析并利于数据库检索；酶自身稳定，不易自降解。用 PMF 检索数据库时，要提高检索准确度，则产生肽谱的酶切点至少要有 2 个，同时相对分子质量大于 5000 Da 的肽片段对检索没有很大帮助。胰蛋白酶和 Glu-C 蛋白酶较多用于肽谱制备，胰蛋白酶水解位点在赖氨酸（K）和精氨酸（R）的羧基端，Glu-C 蛋白酶在 pH 4.0 条件下水解天冬氨酸（D）和谷氨酸（E）的羧基端。MALDI-TOF-MS 适宜测量分子质量在 500Da 以上的分子，在几千道尔顿以内质量准确度较高。

二、凝胶电泳分离蛋白质 PMF 鉴定方法

我们用标准蛋白建立了 PMF 鉴定凝胶电泳分离蛋白质的方法[13]，又在人肺巨细胞癌细胞蛋白和白血病细胞 HL-60 的鉴定中得到应用[14~17]，凝胶电泳分离蛋白 PMF 鉴定的具体步骤如下：

a) 考马斯亮蓝染色凝胶的脱色：将电泳及染色后凝胶中的蛋白带切下，用干净的刀片将胶带切成 1～2 mm^2 大小的胶片，用水清洗几次后用含 50%乙腈、25 mmol/L 碳酸氢铵的溶液 50～100 μl 浸泡胶片，振荡 20 min，弃去溶液，重复 1 至 2 遍至胶片中的蓝色褪尽。

b) 银染凝胶的脱色：配 30 mmol/L 铁氰化钾和 100 mmol/L 硫代硫酸钠储备液，临用前 1:1 混合均匀，用此混合液 30～50 μl 浸泡胶片，涡旋混合器振荡至凝胶中的棕黄色褪尽，弃去溶液，胶片用水洗至无淡黄色溶液产生。

c) 酶切：将胶片真空离心干燥 20 min，使其完全脱水体积缩小，加入 5～10 μl 酶液（0.01 μg/μl，含 25 mmol/L NH_4HCO_3），4℃冰箱中放置 20～30 min，待酶液完全被吸收，补充 5～10 μl 25 mmol/L NH_4HCO_3 液，37 ℃保温 15 h。

d) 肽提取：加 5%TFA 50～100 μl 于 40 ℃保温 1 h，吸出上清液，加 2.5% TFA、50% ACN 50～100 μL 于 30 ℃保温 1 h，吸出上清液，合并上清液，冰冻干燥。加 5～10 μl 0.5%TFA 溶液混匀用于 MALDI-TOF-MS 质谱分析。

e) 数据库检索鉴定蛋白质。

主要使用国际互联网上的免费蛋白数据库检索程序，网址见表 6-2。

表 6-2 PMF 检索程序网址

程序名称	网 址	服务网站
PeptIdent	http://www.expasy.ch/tools/peptident.html	ExPASy
MS-Fit	http://prospector.ucsf.edu/ucsfhtml3.2/msfit.htm	UCSF Mass Spectrometry Facility
ProFound	http://prowl.rockefeller.edu/cgi-bin/ProFound	ProteoMetrics and Rockefeller University
PeptideSearch	http://www.mann.embl-heidelberg.de/GroupPages/PageLink/peptidesearch/Services/PeptideSearch/FR_PeptideSearchFormG4.html	EMBL Protein & Peptide Group
Mascot	http://www.matrixscience.com/cgi/index.pl?page=/search_form_select.html	Matrix Science Ltd., London

检索时根据样品及仪器实际情况输入检索参数限制检索范围，主要有以下几条：

数据库（database）；蛋白种属（organism species or organism classfication，OS or OC）；蛋白的等电点（pI）和相对分子质量（Mw）范围；半胱氨酸的质量修饰；肽质量数误差范围（mass tolerance）；未水解的酶切位点数（maximum number of missed cleavage sites，# MC）；最少相符的肽数目（minimum number of peptides required to match）等。根据检索结果判断蛋白质鉴定结果。

得到可信检索结果的前提是所测的肽质量数准确度要高，参数“Mass tolerance”值设置越小，检出的蛋白数越少，目的蛋白的检索得分值与其他蛋白得分值的差距越大。用兔磷酸化酶 B（PHS2_RABIT；pI 6.77，Mw 97 158.22Da）的 8 个肽质量数（偏差 0.5 Da 以内）输入 PeptIdent，检索参数设 Mw 90～110kDa，pI 6～8，mammalia，最少匹配肽数目 5 个。不同 mass tolerance 的检索结果如表 6-3 所示。从中看出 mass tolerance 至少要为 2.0 Da，目的蛋白的得分值与其他蛋白分值才相差较远。

表 6-3 质量准确度对检索结果的影响

mass tolerance（Da）	检索出的蛋白		
	得分	数目	蛋白名称
0.5	1.0	1	PHS2_RABIT
	0.62	1	PHS2_RAT
1.0	1.0	1	PHS2_RABIT
	0.62	1	PHS2_RAT
2.0	1.0	1	PHS2_RABIT

续表

mass tolerance（Da）	检索出的蛋白		
	得分	数目	蛋白名称
	0.75	2	PHS2 _ RAT；PHS2 _ HUMAN
	0.62	7	非 PHS2 蛋白（略）
3.0	1.0	1	PHS2 _ RABIT
	0.88	3	RB _ MOUSE；PHS2 _ RAT；PHS2 _ HUMAN
	0.75～0.62	22	非 PHS2 蛋白（略）
4.0	1.0	2	PHS2 _ RABIT；PHS2 _ HUMAN
	0.88	7	非 PHS2 蛋白（略）
	0.75～0.62	44	非 PHS2 蛋白（略）

三、凝胶电泳分离蛋白质 PMF 鉴定方法实例

图 6-12 是 PMF 鉴定双向电泳分离蛋白质的实例[17]，此蛋白来源于人白血病细胞，经胰蛋白酶原位酶切，MALDI-TOF-MS 分析得到 PMF，经数据库检索鉴定为人丙酮酸激酶（KPY _ HUMAN）。图中标有 * 的峰是在检索时与数据库中丙酮酸激酶肽质量数理论值相符的肽，图 6-13 为检索参数及检索结果。

KPY1 _ HUMAN 蛋白序列如下，用阴影标记的序列为检索所得蛋白肽段质

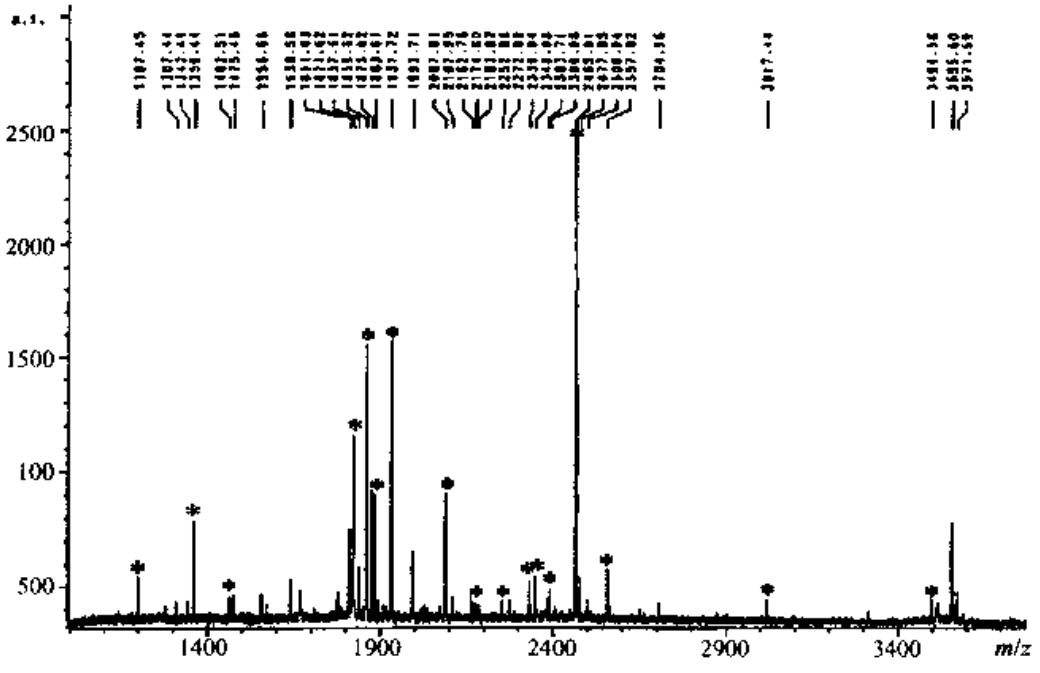


图 6-12　2D 胶上蛋白点的 PMF 图

量数理论值与实验测得 PMF 中质量数相符的肽段序列，质量数相符肽段的序列覆盖率为 50.2%。

Score: 0.51, 18 matching peptides: P14618 (KPYI_HUMAN) pI: 7.57, Mw: 57746.62
PYRUVATE KINASE, M1 ISOZYME (EC 2.7.1.40) (PYRUVATE KINASE MUSCLE ISOZYME) (CYTOSOLIC THYROID HORMONE-BINDING PROTEIN) (CTHBP) (THBP1). - Homo sapiens (Human).

user mass	matching mass	Δ mass (Dalton)	#MC	modification	position	peptide sequence	links
1197.45	1197.582	0.132	0		141-150	ITLDNAYMEK	
1197.45	1197.647	0.1973	0		32-42	LDIDSPPITAR	
1359.44	1359.705	0.2649	0	Cys_CAM: 48	43-55	NTGIICTIGPASR	
1462.51	1462.815	0.3051	0		173-185	IYVDDGLISLQVK	
1821.62	1821.916	0.2963	1		278-293	RFDEILEASDGIMVA R	
1837.61	1837.911	0.3012	1	MSO: 290	278-293	RFDEILEASDGIMVA R	
1859.62	1859.9	0.2796	0		230-245	FGVEQDVDMVFASFI R	
1883.61	1883.903	0.2935	0		73-88	LNFSHGTHEYHAETI K	FindMod
1931.72	1931.986	0.2661	0		383-398	EAEAAIYHLQLFEEL R	
2087.81	2088.087	0.2772	1		383-399	EAEAAIYHLQLFEEL RR	PeptideMass
2174.8	2175.118	0.3179	0		400-421	LAPITSDPTEATAVG AVEASFK	
2252.88	2253.116	0.2359	1		73-91	LNFSHGTHEYHAETI KNVR	BioGraph
2330.94	2331.219	0.279	1		399-421	RLAPITSDPTEATAV GAVEASFK	
2348.98	2349.266	0.2859	1		207-229	GVNLPGAAVDLPAVS EKDIQDLK	
2390.98	2391.26	0.2801	1		504-525	KGDVVIVLIGWRPGS GPTNTMR	
2465.01	2465.292	0.2822	0		92-114	TATESFASDPILYRP VAVALDTK	
2557.02	2557.275	0.2555	1	Cys_CAM: 473	467-488	GIFPVLCKDPVQEAW AEDVDLR	
3017.44	3017.594	0.1541	1		92-119	TATESFASDPILYRP VAVALDTKGPEIR	
3494.58	3494.841	0.0613	1	1xCys_CAM	342-375	AEGSDVANAVLDGAD CIMLSGETAKGDYPL EAVR	

图 6-13　PMF 检索鉴定参数及结果

```
      1          11         21         31         41         51
      |          |          |          |          |          |
  1 SKPHSEAGTA FIQTQQLHAA MADTFLEHMC RLDIDSPPIT ARNTGIICTI GPASRSVETL   60
 61 KEMIKSGMNV ARLNFSHGTH EYHAETIKNV RTATESFASD PILYRPVAVA LDTKGPEIRT  120
121 GLIKGSGTAE VELKKGATLK ITLDNAYMEK CDENILWLDY KNICKVVEVG SKIYVDDGLI  180
181 SLQVKQKGAD FLVTEVENGG SLGSKKGVNL PGAAVDLPAV SEKDIQDLKF GVEQDVDMVF  240
241 ASFIRKASDV HEVRKVLGEK GKNIKIISKI ENHEGVRRFD EILEASDGIM VARGDLGIEI  300
301 PAEKVFLAQK MMIGRCNRAG KPVICATQML ESMIKKPPPT RAEGSDVANA VLDGADCIML  360
361 SGETAKGDYP LEAVRMQHLI AREAEAAIYH LQLFEELRRL APITSDPTEA TAVGAVEASF  420
421 KCCSGAIIVL TKSGRSAHQV ARYRPRAPII AVTRNPQTAR QAHLYRGIFP VLCKDPVQEA  480
481 WAEDVDLRVN FAMNVGKARG FFKKGDVVIV LTGWRPGSGF TNTMRVVPVP
```

用 PMF 方法成功鉴定的关键是所测的肽质量数有足够的准确度，3000 Da 以下的质量数准确度至少要好于 0.5‰。现在较好的 MALDI-TOF-MS 质谱仪在内标校正下已能达到 5 ppm 的质量准确度，使检索鉴定成功率大大提高。但即使这样也不能保证成功鉴定所有蛋白质，有相当一部分蛋白质不能单纯靠 PMF 鉴定，还需要结合序列信息。

PeptIdent
Peptide mass fingerprinting

Name given to unknown protein:	apo21
Species searched:	HOMO SAPIENS (HUMAN)
Database searched:	SWISS-PROT
pI: 8.00	range: 7 9
Mr: 66000	range: 52800 79200
Peptide masses for unknown protein:	1197.45 1307.44 1342.49 1359.44 1462.51 1475.46 1556.66 1638.58 1811.63 1821.62 1837.61 1859.62 1883.61 1931.72 1993.71 2087.81 2107.95 2162.76 2174.80 2183.87 2252.88 2272.89 2330.94 2348.98 2383.71 2390.98 2465.01 2477.05 2500.94 2557.02 2704.96 3017.44 3494.58 3555.60 3571.59
Tolerance:	+0.5 Dalton
Minimum number of peptides required to match:	4
Maximum number of matching proteins to print:	10

Using nonoisotopic masses of the occurring amino acid residues and interpreting your peptide masses as $[H+H]^+$.
Enzyme: Trypsin, allowing for up to 1 missed cleavages (# MC).
Cysteine treated with Iodoacetamide to form carboxyamidomethyl cysteine (Cys _ CAM).
Methionine in oxidized form.

Displaying the first 10 matches.

Score	# peptide matches	AC	ID	Description	pI	Mr
0.51	18	P14618	KPY1 _ HUMAN	PYRUVATE KINASE, M1 ISOZYME (EC 2.7.1.40) (PYRUVATE KINASE MUSCLE ISOZYME) (CYTOSOLIC THYROID HORMONE-BINDING PROTEIN) (CTHBP) (THBP1). -Homo sapiens (Human).	7.57	57746.62
0.51	18	P14786	KPY2 _ HUMAN	PYRUVATE KINASE, M2 ISOZYME (EC 2.7.1.40). -Homo sapiens (Human).	7.95	57782.66
0.17	6	Q15036	Y064 _ HUMAN	HYPOTHETICAL PROTEIN KIAA0064 (HA1355). -Homo sapiens (Human).	7.07	52901.24
0.17	6	Q05038	MYT1 _ HUMAN	MYELIN TRANSCRIPTION FACTOR 1 (MYT1) (MYTI) (PROTEOLIPID PROTEIN BINDING PROTEIN) (PLPBI) (FRAGMENT). -Homo sapiens (Human).	undefined	undefined
0.17	6	P43250	GRK6 _ HUMAN	G PROTEIN-COUPLED RECEPTOR KINASE GRK6 (EC 2.7.1.-) Homo sapiens (Human).	8.33	65968.92
0.17	6	P08684	CP34 _ HUMAN	CYTOCHROME P450 3A4 (EC 1.14.14.1) (CYPIIIA4) (NIFEDIPINE OXIDASE) (NF-25) (P450-PCN1). -Homo sapiens (Human).	8.30	57299.07
0.14	5	O75616	ERAH _ HUMAN	GTP-BINDING PROTEIN ERA HOMOLOG (HERA) (FRAGMENT). -Homo sapiens (Human).	undefined	undefined
0.14	5	Q13939	CALI _ HUMAN	CALICIN (FRAGMENT). -Homo sapiens (Human).	undefined	undefined
0.14	5	P14210	HGF _ HUMAN _ 1	CHAIN 1: HEPATOCYTE GROWTH FACTOR ALPHA CHATN. -Homo sapiens (Human).	7.69	53684.21
0.14	5	P40939	ECHA _ HUMAN _ 1	CHAIN 1: MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA-Homo sapiens (Human).	8.98	78969.76

第四节　串联质谱数据鉴定蛋白质技术
（肽序列标签与蛋白质从头测序技术）

一、串联质谱测定多肽序列原理

蛋白质由20种氨基酸组成，一段3个氨基酸的肽有20^3种可能排列方式，4个氨基酸的肽有20^4种可能排列方式，一个特定序列的4肽出现的概率为1/160 000。所以5、6个氨基酸残基的序列片段在一个蛋白质组中已具有很高的特异性，可以用来鉴定蛋白质。

串联质谱仪可获得肽段序列信息。肽段母离子在质谱仪的碰撞室经高流速惰性气体碰撞解离，沿肽链在酰胺键处断裂并形成子离子。肽键断裂时，会产生a、b、c型和x、y、z型系列离子（图6-14），a、b、c型离子保留肽链N端，电

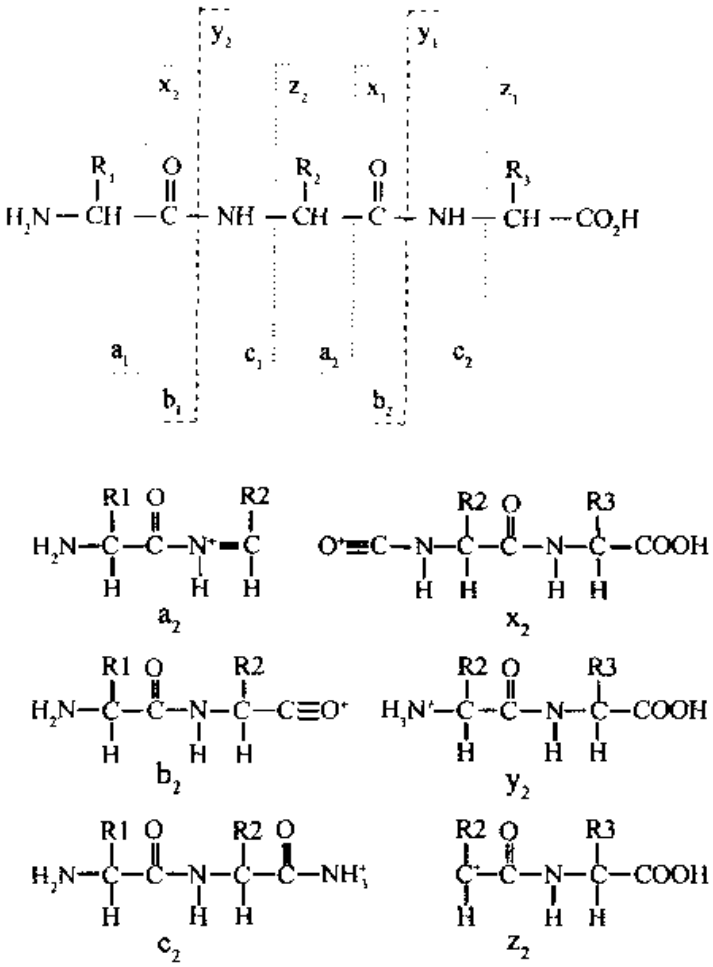


图6-14　多肽串联质谱碎片离子示意图

荷留在离子C端，x、y、z型离子保留肽链C端，电荷留在离子N端，其中b型和y型离子在质谱图中较多见，丰度较高。还会出现b-H_2O和y-NH_3等离子形式。y、b系列相邻离子的质量差，即为氨基酸残基质量，根据完整或互补的y、b系列离子可推算出氨基酸的序列，20种基本氨基酸残基的相对分子质量见表6-4。质谱方法不能分辨亮氨酸（Leu，L）与异亮氨酸（Ile，I），二者为同分异构体，残基相对分子质量均为113.08406（momo）。谷氨酰胺（Gln，Q）与赖氨酸（Lys，K）的残基相对分子质量十分接近，分别为128.05858（mono）与128.09496（mono），需要质谱仪的质量分辨率和准确度十分高才能正确分析。

表6-4　20种氨基酸残基（ NH　HCR—C═O ）相对分子质量

氨基酸名称	单同位素相对分子质量 *monoisotopic*	平均相对分子质量 average
甘氨酸 Glycine (G)	57.02146	57.0519
丙氨酸 Alanine (A)	71.03711	71.0788
丝氨酸 Serine (S)	87.03203	87.0782
脯氨酸 Proline (P)	97.05276	97.1167
缬氨酸 Valine (V)	99.06841	99.1326
苏氨酸 Threonine (T)	101.04768	101.1051
半胱氨酸 Cysteine (C)	103.00919	103.1388
异亮氨酸 Isoleucine (I)	113.08406	113.1594
亮氨酸 Leucine (L)	113.08406	113.1594
天冬酰胺 Asparagine (N)	114.04293	114.1038
天冬氨酸 Aspartic acid (D)	115.02694	115.0886
谷氨酰胺 Glutamine (Q)	128.05858	128.1307
赖氨酸 Lysine (K)	128.09496	128.1741
谷氨酸 Glutamic acid (E)	129.04259	129.1155
甲硫氨酸 Methionine (M)	131.04049	131.1926
组氨酸 Histidine (H)	137.05891	137.1411
苯丙氨酸 Phenylalanine (F)	147.06841	147.1766
精氨酸 Arginine (R)	156.10111	156.1875
酪氨酸 Tyrosine (Y)	163.06333	163.1760
色氨酸 Tryptophan (W)	186.07931	186.2132

二、串联质谱分析肽段序列实例

我们曾经分析过核糖核酸酶 B 的一级结构[19]，核糖核酸酶 B 用糖苷内切酶除糖后，用 Glu-C 酶切，在酶切产物的肽指纹谱中，有一段 m/z 4792.23 的肽段，其质量数与任何理论肽段质量数均不相符。后来经串联质谱分析，测得其部分序列为 TSAASSSNYCNQM，与理论序列比较后发现，此序列所在的酶切肽段理论值与实测值相差 203，恰好是一个 N-乙酰氨基己糖的质量数，表明此肽段上还连接有一个 N-乙酰氨基己糖。图 6-15 是 m/z 4792.23 的串联质谱图及解析出的部分序列。

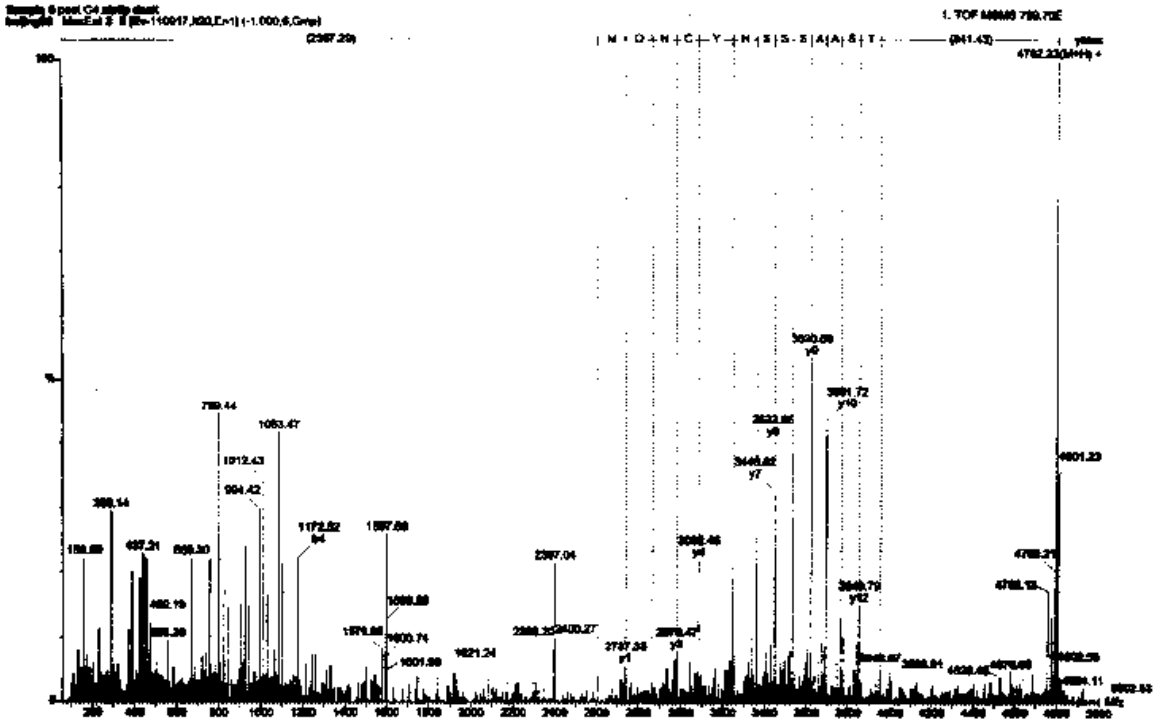


图 6-15　核糖核酸酶 B 的 Glu-C 酶切肽段 m/z 4792.23 的串联质谱图及解析出的序列

肽段理论序列：

RQHMDSS [TSAASSSNYCNQM] MKSR [NLT] KDRCKPVNTFVHE

阴影标记的部分为串联质谱解析出的序列，方框内序列为天冬酰胺序列，N 为糖基化位点。

肽段理论质量 $[M+H]^+$ m/z：4589.10

肽段实测质量 $[M+H]^+$ m/z：4792.23

实测值与理论值之差：4792.23 − 4589.10 = 203.13（*N*-乙酰氨基己糖质量数）

对所用的糖苷内切酶研究后发现，此酶为两种酶的混合物，在本实验所用的酸性条件下，发挥作用的糖苷内切酶的切点在第二个糖苷键，所以第一个 N-乙酰氨基葡萄糖仍然留在天冬酰胺残基上，造成此肽段质量数理论值与实测值的偏差。

三、肽序列标签鉴定技术

与 PMF 谱图相比，串联质谱技术获取的肽序列谱图相对较复杂，需要计算软件帮助计算识别 b、y 等各系列离子，从谱图读出完整序列的解析较困难。数据库检索鉴定蛋白质时，可用读出的部分氨基酸序列结合此段序列前后的离子质量和肽段母离子质量，在数据库中查寻，这一鉴定方法称为肽序列标签技术[20] (peptide sequence tag，PST)，或者直接用串联质谱数据进行检索。常用的串联质谱数据鉴定蛋白质检索程序网站见表 6-5。

表 6-5　串联质谱数据检索程序网址

程序名称	网　址	服务网站
MS-Tag	http://prospector.ucsf.edu/ucsfhtml3.2/mstagfd.htm http://prospector.ucsf.edu/ucsfhtml3.2/mstag.htm	UCSF Mass Spectrometry Facility
MS-Seq	http://prospector.ucsf.edu/ucsfhtml3.2/msseq.htm	
PepFrag	http://prowl1.rockefeller.edu/prowl/pepfragch.html	ProteoMetrics and Rockefeller University
MOWSE	http://srs.hgmp.mrc.ac.uk/cgi-bin/mowse	The UK Human Genome Mapping Project Resource Centre
Mascot	http://www.matrixscience.com/cgi/index.pl?page=/search_form_select.html	Matrix Science Ltd., London
PeptideSearch	http://www.mann.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html	EMBL Protein & Peptide Group

PST 检索鉴定实例：

例 1：从马心肌红蛋白的胰蛋白酶酶切肽段中选取母离子进行串联质谱分析及蛋白质鉴定。图 6-16 为马心肌红蛋白的胰蛋白酶酶切产物的 ESI-Q-TOF 肽指纹图，选取双电荷离子峰 m/z 908.4 为母离子，图 6-17 为此母离子的串联质谱图，根据质谱数据解析出肽序列标签(390.3)VNLVQQW(1257.8)，用于数据库检索查寻。检索程序为 PeptideSearch，检索参数及结果见图 6-18。

检索程序正确地查寻出可产生这一肽序列标签的肽段序列，结果还显示这一序列只出现在肌红蛋白的序列中。

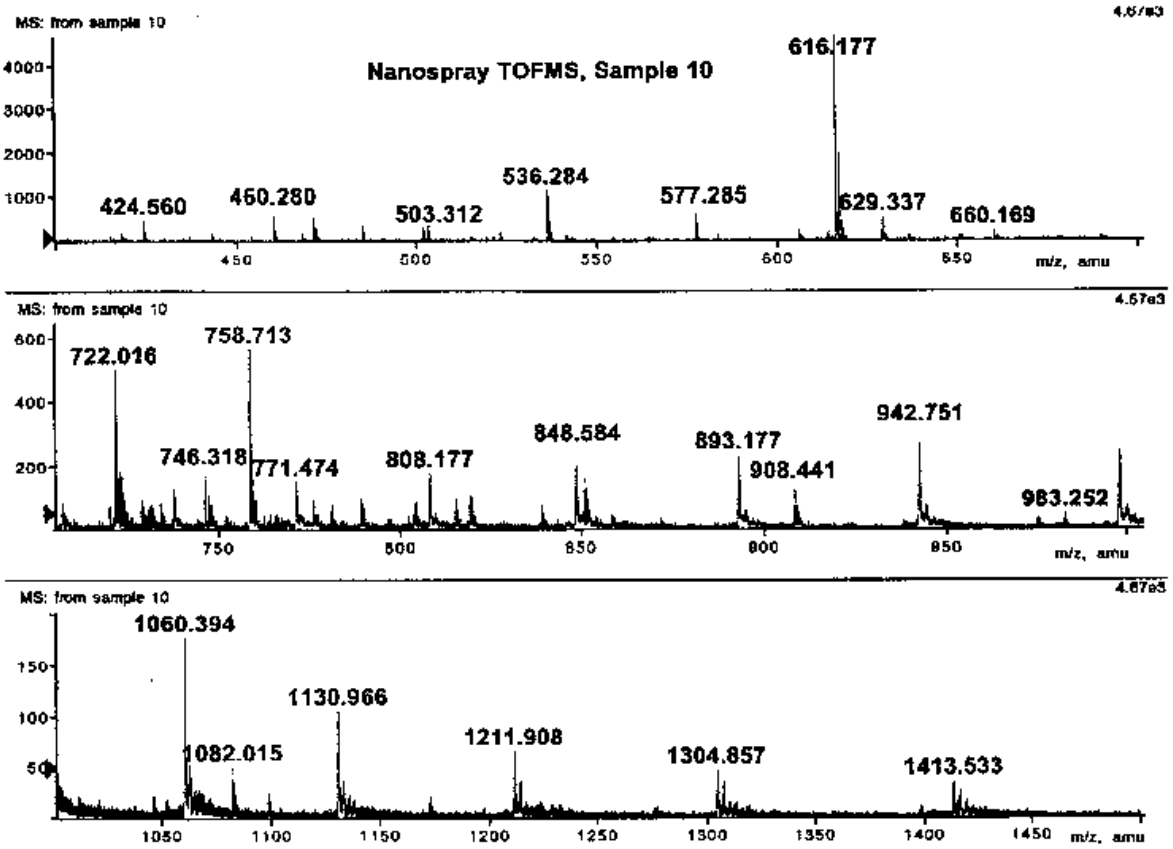


图 6 16　马心肌红蛋白胰蛋白酶酶切产物 ESI-Q-TOF 肽指纹图

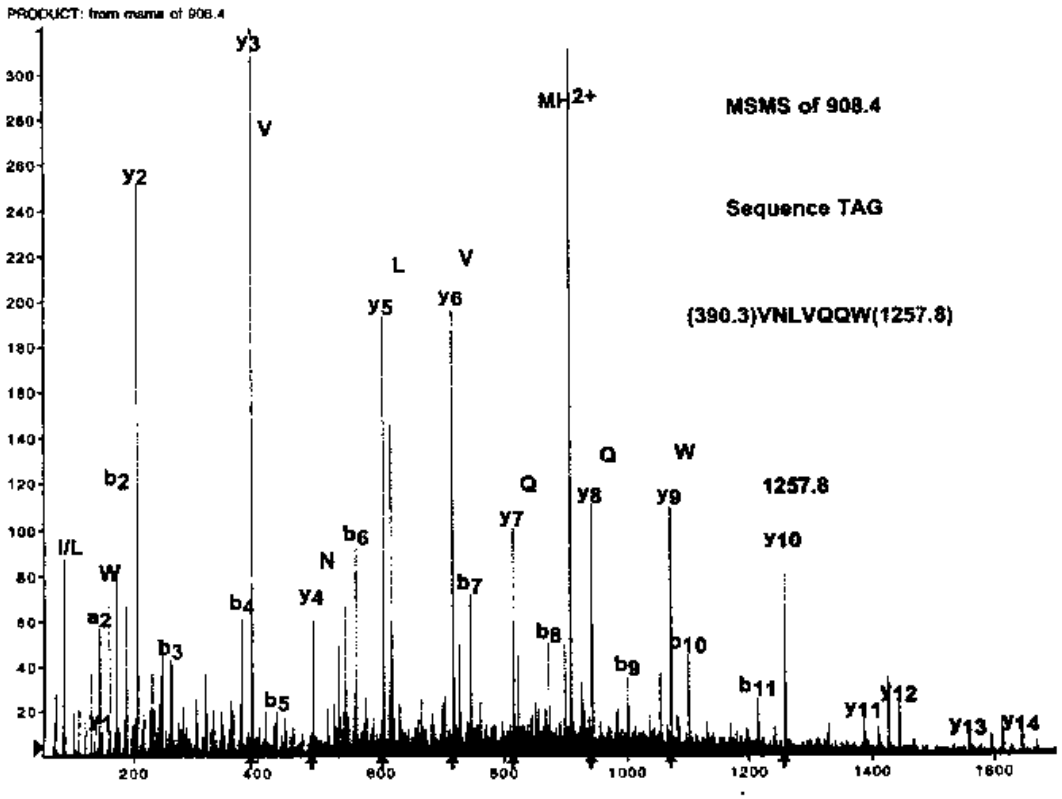


图 6-17　马心肌红蛋白胰蛋白酶酶切肽段 m/z 908.4 的串联质谱图

Search parameters

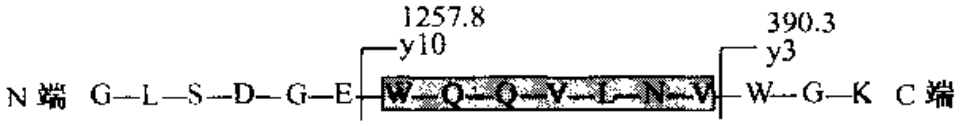
Sequence tag	(390.3)vnlvqqw(1257.8)
Protein mass range	0-300 kDa
Cleavage agent	Trypsin
Peptide mass accuracy	0.5 Da
Methionine is	Native
Cysteine is	Cys
Peptide mass	1814.8
Match regions	1 and 2 and 3
Search by	Y-type sequence ions
Allowed number of errors	1
Nominal mass	Isoleucine equals Leucine
Nominal mass	Glutamine equals Lysine
N terminal specific	No
C terminal specific	No

Search result

6 matches were found. Showing matches 1 through 6.

Peptide Sequence matched/ Peptide found	Mass [kDa]	Database accession	Protein Name
GLSDGEWQQVLNVWGK	16.95	swissnew:P02188	P02188|MYG_HORSE MYOGLOBIN.//:p
GLSDGEWQQVLNVWGK	16.92	pdb:1BJE (not known by SRS)	1BJE H64T VARIANT OF MYOGLOBI
GLSDGEWQQVLNVWGK	16.98	pdb:1HRM (not known by SRS)	1HRM MYOGLOBIN MUTANT WITH HI
GLSDGEWQQVLNVWGK	16.98	pdb:1RSE (not known by SRS)	1RSE MYOGLOBIN (HORSE HEART)
GLSDGEWQQVLNVWGK	16.95	pdb:1XCH (not known by SRS)	1XCH MYOGLOBIN (HORSE HEART)
GLSDGEWQQVLNVWGK	16.98	pdb:1YMA (not known by SRS)	1YMA MYOGLOBIN (HORSE HEART)

图 6-18　PST(390.3)VNLVQQW(1257.8)的检索参数及结果



例 2:图 6-19 为一母离子质荷比为 m/z 740.4(双电荷离子,中性分子质量为 1478.8Da)的肽段的串联质谱测序图,检索用肽序列标签为"(685.43)QFG(1017.6)",图 6-20 是 PST 检索参数及鉴定结果。

检索结果表明,在所设参数范围内,检索程序在蛋白质数据库中检索到一段肽序列"LGEYGFQNALLVR",该序列可以产生肽序列标签"(685.43)QFG(1017.6)",该序列肽段来自牛血清白蛋白。

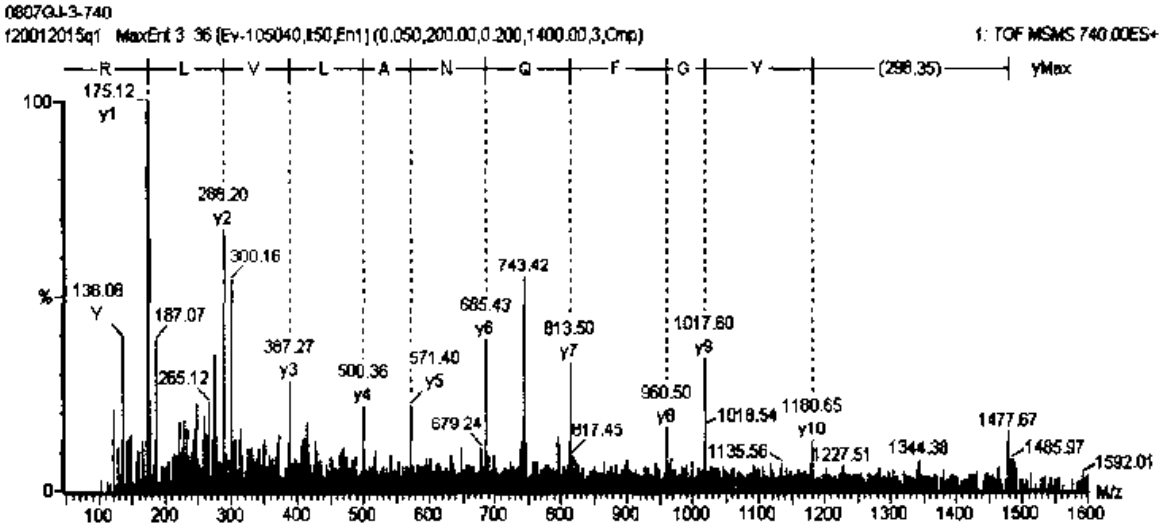


图 6-19 MS/MS 串联质谱图

四、PSD 质谱数据检索鉴定技术

MALDI-RETOF 质谱仪做序列分析得到的是肽段源后衰变(PSD)产生的亚稳离子峰,PSD 数据不宜形成完整的 y 或 b 离子系列,直接解析序列较困难,有些检索程序可以直接利用不经处理的 PSD 谱图数据进行数据库检索查寻鉴定[21]。

例 1:图 6-21 是血管紧张素肽(Angiotensin)的 PSD 质谱图,质谱数据直接进行联机检索,检索结果如下:

根据 PSD 质谱数据,检索程序自动选取 38 个丰度较高的谱峰质量数,用这些质量数在数据库中查寻可能产生相似 PSD 数据的肽段,结果正确检索出血管紧张素肽(angiotensin)的序列。PSD 数据中共有 19 个离子质量数与血管紧张素肽的

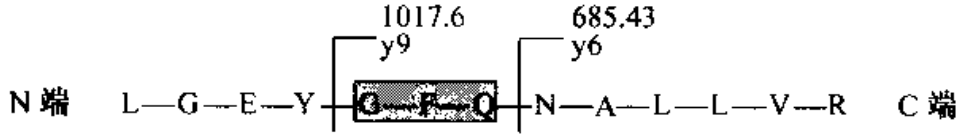
Search parameters

Sequence tag	(685.43)qEg(1017.6)
Protein mass range	0-300 kDa
Cleavage agent	Trypsin
Peptide mass accuracy	0.5 Da
Methionine is	Native
Cysteine is	Cys
Peptide mass	1478.8
Match regions	1 and 2 and 3
Search by	Y-type sequence ions
Allowed number of errors	0
Nominal mass	Isoleucine equals Leucine
Nominal mass	Glutamine equals Lysine
N terminal specific	Yes
C terminal specific	Yes

Search result

Peptide Sequence matched/ Peptide found	Mass [kDa]	Database accession	Protein Name
LGEYGFQNALLVR	68.61	swissnew:P49822	ALBU_CANFA SERUM ALBUMIN PRECUR
LGEYGFQNALIVR	69.29	swissprot:P02769	ALBU_BOVIN SERUM ALBUMIN PRECUR
LGEYGFQNALLVR	32.64	swissprot:P49822	ALBU_CANFA SERUM ALBUMIN (FRAGM
LGEYGFQNALLVR	68.66	swissprot:P49064	ALBU_FELCA SERUM ALBUMIN PRECUR

图 6-20　Peptide Sequence Tag 检索结果



理论值相符，检索结果列出了血管紧张素肽的 PSD 数据理论值及各种离子的类型，并标记出与实验测定值相符的质量。

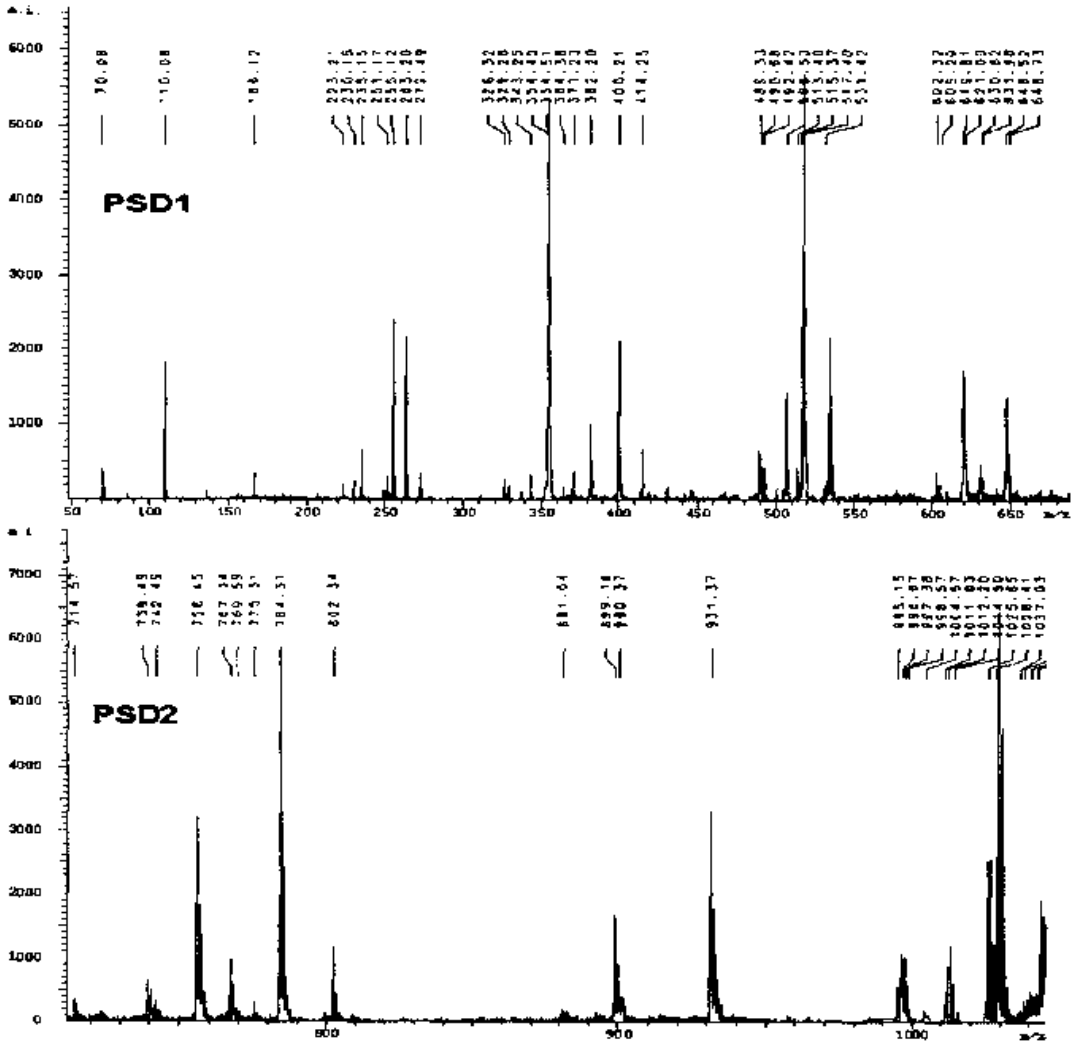


图 6-21　血管紧张素肽(m/z1046.54, +1)的 PSD 质谱图

例 2:图 6-22 是双向电泳分离的 α 粒子辐射诱导人支气管上皮细胞中的一个蛋白质点的 MALDI-TOF-MS 肽指纹谱,选取其中 m/z 1716.05 的峰进行 PSD 分析,图 6-23 是 PSD 的谱图,用该图数据在 Mascot 网站检索,检索出该蛋白质为 maspin precursor,检索结果见图 6-24。

{MATRIX} {SCIENCE} Mascot Search Results

Peptide View

MS/MS Fragmentation of DRVYIHPF
Found in ANG2_BOTJA, ANGIOTENSIN-LIKE PEPTIDE II (FRAGMENT) - BOTHROPS JARARACA (JARARACA)

Match to Query 1 (1046.54,1+)
From data file DATA.TXT

Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 0 to 1050 Da

Monoisotopic mass of neutral peptide (Mr): 1045.53

Ions Score: 33 Matches (Bold): 19/42 fragment ions using 38 most intense peaks

#:	1	2	3	4	5	6	7	8
a:	88.04	244.14	**343.21**	**506.27**	**619.36**	**756.42**	853.47	1000.54
a-NH_3:	71.01	227.11	326.18	**489.25**	602.33	**739.39**	836.44	983.51
b:	116.03	**272.14**	371.20	534.27	647.35	**784.41**	881.46	**1028.53**
b-NH_3:	99.01	**255.11**	**354.18**	**517.24**	**630.33**	**767.38**	864.44	1011.51
Seq:	D	R	V	Y	I	H	P	F
y-NH_3:	1029.52	914.49	758.39	659.32	496.26	383.17	246.11	149.06
y:	1046.54	**931.52**	775.41	676.35	**513.28**	**400.20**	**263.14**	**166.09**
#:	8	7	6	5	4	3	2	1

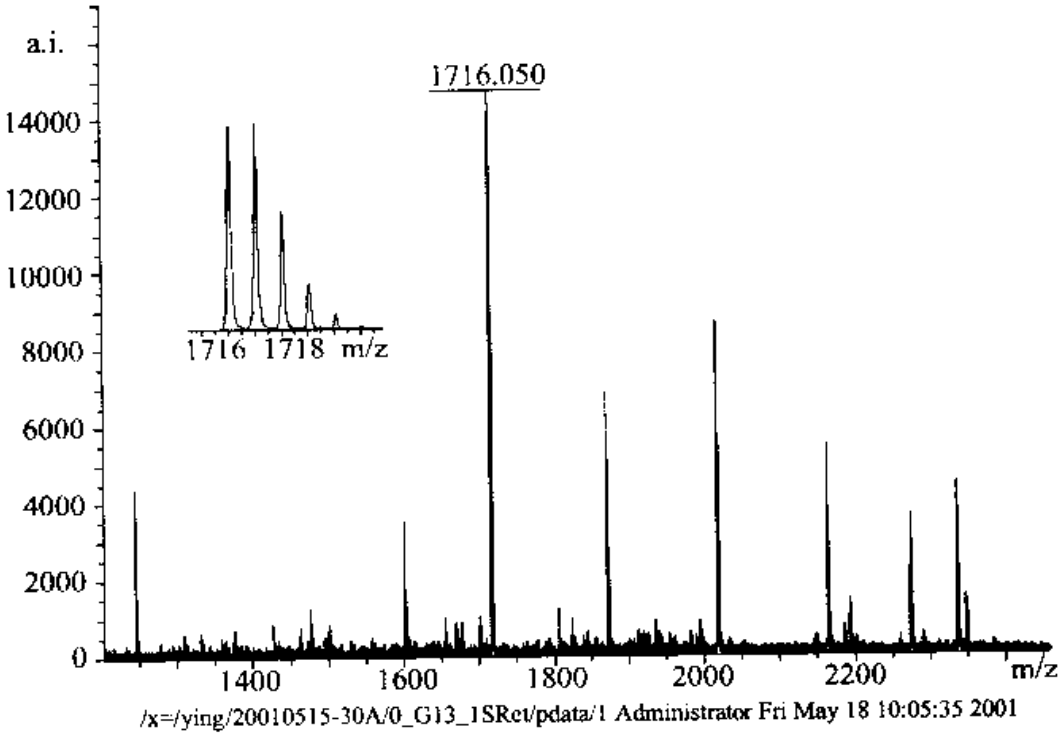


图 6-22　α 粒子辐射诱导人支气管上皮细胞中蛋白质点肽质量指纹谱

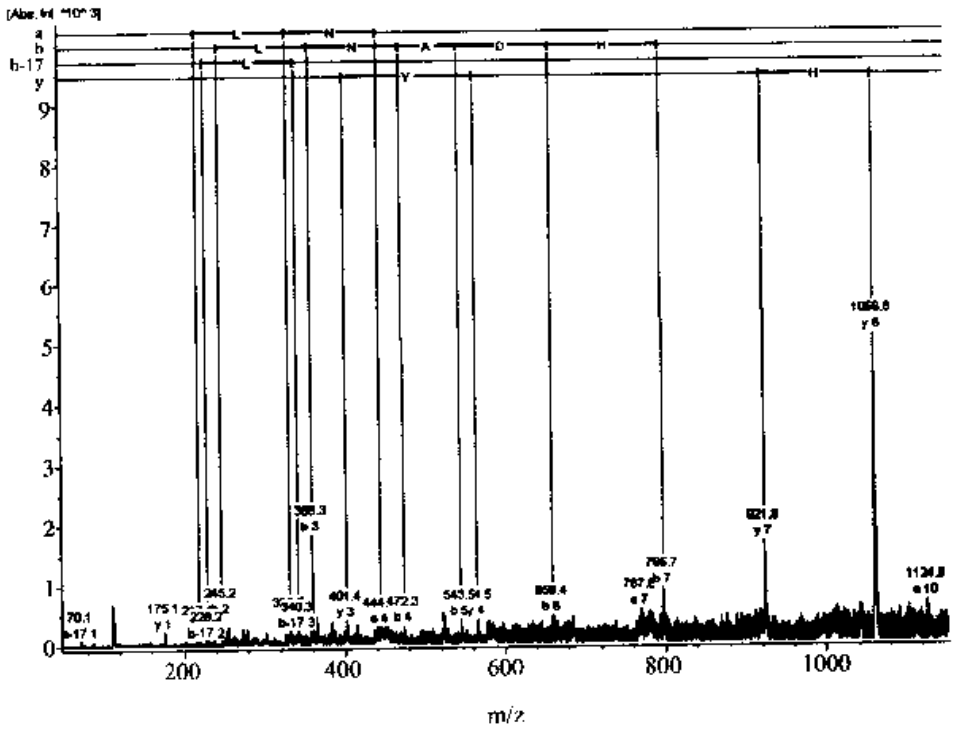


图 6-23　m/z 1716.05 的峰的 PSD 谱图

MATRIX SCIENCE **Mascot Search Results**

Peptide View

MS/MS Fragmentation of DELNADHPFIYHR

Found in MASP_ HUMAN, MASPIN PRECURSOR (PROTEASE INHIBITOR 5). -HOMO SAPIENS (HUMAN).

Match to Query 1(1716.95, 1+)

From data file DATA TXT

Average mass of neutral peptide (Mr): 1715.93

Fixed modifications: Carbamidomethyl (C)

Ions Score: 45 Matches (Bold Red): 17/78 fragment ions using 32 most intense peaks

#:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
a:	88.09	217.20	330.36	444.47	515.54	630.63	767.77	864.89	1012.07	1125.23	1288.40	1401.56	1514.72	1670.91
a*:	71.06	200.17	313.33	427.43	498.51	613.60	750.74	847.86	995.04	1108.20	1271.37	1384.53	1497.69	1653.88
b:	116.10	245.21	358.37	472.48	543.55	658.64	795.78	892.90	1040.08	1153.24	1316.41	1429.57	1542.73	1698.92
b*:	99.07	228.18	341.34	455.44	526.52	641.61	778.75	875.87	1023.05	1136.21	1299.38	1412.54	1525.70	1681.89
Seq:	D	E	L	N	A	D	H	P	F	I	Y	I	I	R
y*:	1699.90	1584.82	1455.70	1342.54	1228.44	1157.36	1042.27	905.13	808.01	660.83	547.68	384.50	271.34	158.18
y:	1716.93	1601.85	1472.73	1359.57	1245.47	1174.39	1059.30	922.16	825.04	677.87	564.71	401.53	288.37	175.21
#:	14	13	12	11	10	9	8	7	6	5	4	3	2	1
Immonium ions:														
Imm:	88.09	102.11	86.16	87.10	44.08	88.09	110.14	70.11	120.17	86.16	136.17	86.16	86.16	129.19

图 6-24 Mascot 网站检索结果

五、^{18}O 标记从头测序技术

串联质谱图中谱峰的类型较多，峰的数目也很多，如果样品的纯度不是很好，样品溶液中有盐干扰，谱图噪声较大，要从中找出完整的 y 或 b 型离子系列，读出完整的肽序列较困难。PST 等检索方式虽然提供了用部分串联质谱数据鉴定蛋白质的途径，但对于结构未知的新蛋白，如果要设计 DNA 探针或在蛋白质数据库及表达序列标签库(EST)中进行同源性检索，至少需要一段含 10～15 个氨基酸残基的序列。

为了提高对串联质谱数据的解读能力，尤其是对 y 或 b 型离子系列的辨别能力，实现对肽段的从头测序(de novo sequencing)，可利用化学修饰的方法使该种离子具有某种特征，从而易于辨别。目前较常用的方法是^{18}O 标记法[22]。具体做法是，蛋白质在酶解时，酶解液中 $H_2{}^{18}O$ 和 $H_2{}^{16}O$ 各占 50%，这样酶解后肽段的羧基端上的羟基氧的$^{18}O/^{16}O$ 同位素丰度比为 1∶1，使得 y 型离子系列(均含有肽链的游离羧基端)的 M+2/M 的同位素峰的丰度比高于正常未标记^{18}O 离子的同位素峰的丰度比(图 6-25)，根据这一特征可以较容易地分辨出 y 系列离子，用这种方法可以较准确地读出 10 个以上的氨基酸序列。

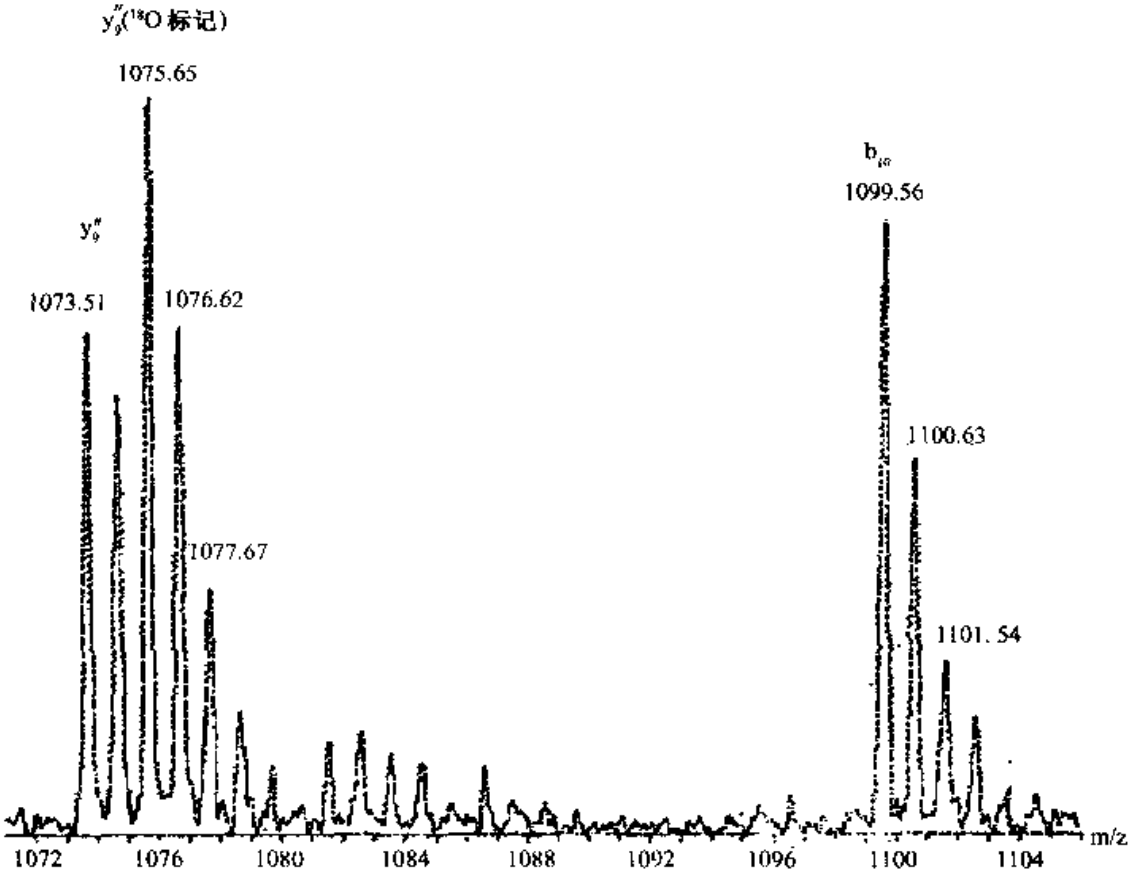


图 6-25 ¹⁸O 标记 y 型离子

参考文献

1 应万涛，钱小红. 生物质谱技术应用及进展. 军事医学科学院院刊，2000，24(2)：146～150

2 万晶宏，贺福初. 蛋白质组技术的研究进展. 科学通报，1999，44(9)：904～910

3 成海平，钱小红. 蛋白质组研究的技术体系及其进展. 生物化学与生物物理进展，2000，27(6)：584～588

4 Karas M，Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10 000 daltons. Anal Chem，1988，60(20)：2299～2301

5 Whitehouse C M，Dreyer R N，Yamashita M，Fenn J B. Electrospray interface for liquid chromatographs and mass spectrometers. Anal Chem，1985，57(3)：659～675

6 Fenn J B，Mann M，Meng C K，Wong S F，Whitehouse C M. Electrospray ionization for mass spectrometry of large biomolecules. Science，1989，246(4926)：64～71

7 Barbara S L，Charles N M. eds. Mass spectromtry of biological materials. New York：Marcel Dekker，2nd ed，1998

8 Wilkins M R，Williams K L，Appel R D，Hochstrasser D F eds. Proteome Research：New frontiers in functional genomics. Berlin：Springer-Verlag，1997

9 Roepstorff P. Mass spectrometry in protein studies from genome to function. Current Opinion in Biotechnolo-

gy, 1997, 8:6～13

10 钱小红. 蛋白质组与生物质谱技术. 质谱学报, 1998, 19(4):53～58

11 Mann M, Hojrup P, Roepstorff P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biol Mass Spectrom, 1993, 22(6):338～345

12 Yates J R 3rd, Speicher S, Griffin P R, Hunkapiller T. Peptide mass maps: a highly informative approach to protein identification. Anal Biochem, 1993, 214(2):397～408

13 王京兰, 蔡耘, 王杰, 董方霆, 贺福初, 钱小红. 蛋白质组研究中肽质量指纹图制备与鉴定方法初探. 军事医学科学院院刊, 1999, 23(1):71～72

14 王京兰, 万晶宏, 罗凌, 熊少祥, 王光辉, 贺福初, 钱小红. 蛋白质组研究中肽质量指纹谱鉴定方法的建立及应用. 生物化学与生物物理学报, 2000, 32(4):373～378

15 Xiaohong QIAN, Jinghong WAN, Jinglan WANG et al. Identification of proteins from human lung cancer isolated by two-dimensional electrophoresis by MALDI-TOF-MS and N-terminal sequencing. ICES'99 program and abstracts (Japanese Journal of Electrophoresis), 1999, 43:28

16 钱小红, 王京兰, 蔡耘等. 肽质量指纹谱技术的建立及其在人肺癌细胞蛋白质组研究中的初步应用. 1999(首届国际)质谱学术报告会论文摘要 (质谱学报 1999, 20(3－4):75)

17 Jinghong WAN, Jinglan WANG, Haiping Cheng, Yongtao Yu, Guichun Xing, Zongyin Qiu, Xiaohong Qian, Fuchu He. Proteomic analysis of apoptosis initiation induced by all-trans retinoic acid in human acute promyelocytic leukemia cells. Electrophoresis, 2001, 22:3026～3037

18 Xiaohong QIAN, Jinghong WAN, Jinglan WANG *et al*. Proteome study of Leukemia cell apoptosis related proteins by two-dimensional gel electrophoresis and MALDI-TOF-MS. ASMS, 2000

19 蔡耘, 王京兰, 邱丰和, 钱小红. MALDI-TOF-MS 研究糖蛋白核糖核酸酶 B 的一级结构. 质谱学报, 2000, 21(1):24～28

20 Mann M, Wilm M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Anal Chem, 1994, 66(24):4390～4399

21 Griffin P R, MacCoss M J, Eng J K, Blevins R A, Aaronson J S, Yates J R 3rd. Direct database searching with MALDI-PSD spectra of peptides. Rapid Commun Mass Spectrom, 1995, 9(15):1546～1551

22 Shevchenko A, Chernushevich-I, Ens W, Standing KG, Thomson B, Wilm M, Mann M. Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. Rapid-Commun-Mass-Spectrom, 1997, 11(9): 1015～1024

(王京兰)

第七章　蛋白质翻译后修饰的鉴定

出于功能的需要，许多蛋白质在翻译中或翻译后会在氨基酸链上共价结合各种非肽类基团，形成翻译后修饰，修饰基团约有 20 多种类型，常见的如磷酸化和糖基化修饰。有些修饰基团只出现在蛋白的 N 端，与氨基酸种类无关；有些却只出现在某几个氨基酸残基上，与氨基酸位置无关；还有些修饰现象既与氨基酸种类有关，又与其位置有关。这些修饰基团会影响蛋白质的相对分子质量和等电点，使其在双向电泳胶上偏离其理论位置。蛋白质的翻译后修饰与其活性与功能状态有关，也与蛋白质所在细胞的种类和生命周期相关，因此翻译后修饰也是蛋白质鉴定工作的一项重要任务。由于蛋白质组学的研究策略是从整体上了解细胞蛋白的功能状态，所以从整体上了解蛋白质的翻译后修饰也是蛋白质组研究正在面临的技术挑战。蛋白质的糖基化修饰和磷酸化修饰在生命活动中具有重要的调控作用，因此也是目前蛋白质组中翻译后修饰研究的热点。

第一节　磷酸化蛋白质的鉴定

一、生物体内的蛋白质磷酸化修饰及其功能

生物体能迅速对体内环境变化和外界环境刺激产生应答反应，这些反应过程靠复杂的调控机制调节，其中大多数调控机制是由蛋白质的构象变化所介导的。而蛋白质本身的构象变化常常是通过变构效应和蛋白质一级结构上发生的各种共价修饰来实现的，如二硫键的配对、蛋白水解酶的加工、糖基化和磷酸化修饰等，而蛋白质磷酸化是最常见、最重要的共价修饰方式。在哺乳动物细胞周期中，大约有 1/3 的蛋白质发生过磷酸化的修饰，在脊椎动物基因组中，有 5% 的基因编码的蛋白质是参与磷酸化和去磷酸化过程的蛋白激酶（kinase）和磷酸（酯）酶（phosphatase）[1]。蛋白质的磷酸化和去磷酸化这一可逆过程几乎调节着生命活动的所有过程，包括细胞的增殖、发育和分化、细胞骨架调控、细胞凋亡、神经活动、肌肉收缩、新陈代谢、肿瘤发生等，并且可逆的蛋白质磷酸化是目前所知道的最主要的信号转导方式。蛋白质磷酸化过程中磷酸基团的来源是细胞内主要的能量分子 ATP，ATP 在细胞内很丰富，可以确保磷酸化修饰有效地进行。在蛋白激酶的作用下蛋白质被磷酸化，在磷酸酶的作用下蛋白质被去磷酸化，磷酸化与去磷酸化是一个可逆的平衡反应，蛋白激酶和磷酸酶二者的作用平衡决定了蛋白质的磷酸化

状态。被磷酸化修饰的蛋白质可以改变它所带的电荷，也可以改变催化基团的性质，或者改变蛋白质的构象，或者改变蛋白质的亚细胞分布，表现为蛋白质功能变化的多样性[2]。

真核细胞蛋白质的磷酸化翻译后修饰一般是O-磷酸化，即磷酸化位点在蛋白质的丝氨酸(Ser)、苏氨酸(Thr)和酪氨酸(Tyr)残基侧链的羟基上，而原核生物中蛋白质的磷酸化位点常在组氨酸(His)、天冬氨酸(Asp)和谷氨酸(Glu)，此外，在赖氨酸(Lys)、精氨酸(Arg)和半胱氨酸(Cys)上也会发生磷酸化修饰。不同的蛋白激酶可识别和修饰不同蛋白质的不同位点(不同的氨基酸顺序)。图7-1是真核细胞蛋白质中常见的几种磷酸化修饰氨基酸残基的结构

胶电泳技术。细胞蛋白质经二维凝胶电泳分离后，磷酸化蛋白质的检测通常用^{32}P放射性标记放射自显影和抗磷酸氨基酸抗体免疫反应的方法，磷酸化蛋白质的鉴定可以用肽质量指纹谱方法或串联质谱测序方法，磷酸化位点的鉴定可用磷酸酶法、串联质谱测序法等，有些情况下，在质谱分析前还需要对磷酸肽进行富集，如金属螯合离子亲和提取法。在本章的下面部分将分别对这些方法予以介绍。

三、磷酸化蛋白质的检测

1. ^{32}P 放射性标记法

从蛋白质组水平检测磷酸化蛋白质最经典的方法是^{32}P放射性标记法。也就是用放射性^{32}P标记的ATP处理细胞，使^{32}P掺入磷酸化蛋白质。具体方法是[4]：培养待标记的细胞至适当的生长期，在生长旺盛的细胞中磷酸盐的转运达到最高峰，用不含磷酸盐的标记培养液(如DMEM)替换原来的细胞培养液，并加入[^{32}P]标记磷酸盐共培养，使细胞ATP库与^{32}P平衡，蛋白激酶利用被放射标记的ATP使底物磷酸化。培养约2 h后裂解细胞提取蛋白质，随后进行2-D电泳分离，放射自显影检测磷酸化蛋白质。

此方法的局限性在于，对某些磷酸转换速率较低的磷酸化蛋白质，只能掺入少量的放射性磷酸盐，有可能检测不到。另外，不同的氨基酸有不一样的磷酸化/去磷酸化代谢速率，所以掺入[^{32}P]磷酸盐的速率也是不一样的，在不同磷酸化氨基酸之间进行定量比较时要注意这一问题[1]。放射性标记方法虽然灵敏而且直观，但是因为存在放射性污染的问题，所以在相当多的实验室是不适用的。

2. 抗体免疫印迹检测法

用抗磷酸氨基酸抗体与磷酸化蛋白质进行免疫印迹反应(western blotting)来检出磷酸化蛋白质也是目前较常用的磷酸化蛋白质分析方法。现在已经有商品化的抗酪氨酸、丝氨酸和苏氨酸磷酸盐抗体，其中抗酪氨酸磷酸盐抗体的特异性最好，也最为常用。抗磷酸化丝氨酸和苏氨酸抗体的特异性不太高，因为抗原决定簇太小，抗原抗体结合位点有空间障碍，所以某些磷酸化蛋白质不能被检测。以下是用抗体免疫印迹检测法分析的一般步骤[1,4]。

a) 细胞蛋白质样品的制备

细胞培养结束后加裂解液提取细胞蛋白质，裂解液中要加入一定浓度的酶抑制剂，以避免细胞裂解后发生的蛋白质磷酸化或去磷酸化作用。加入氟化钠可抑制丝氨酸/苏氨酸磷酸酶的作用，加入钒酸钠可抑制酪氨酸磷酸酶的作用，加入EDTA可抑制蛋白激酶的活性。由于许多参与信号转导的磷酸化蛋白质都是跨膜蛋白质，在细胞裂解液中加入硫脲会增加这些膜蛋白和核蛋白在一向IPG胶上的溶解性。

b) 一维或二维电泳分离细胞蛋白质

按照常规 SDS-PAGE 或双向电泳的操作步骤进行蛋白质分离。

c) 转膜

电泳结束后将胶上蛋白质电转印到膜上，常用的转印膜有硝酸纤维素膜(nitrocellulose)和 PVDF 膜(polyvinylidene fluorine，聚偏二氟乙烯)。转印的效率取决于蛋白质从胶上转移到膜上的效率以及蛋白质与膜结合的效率，转印缓冲液中含有一定浓度的甲醇(10%～20%)可以提高蛋白质与印迹膜的结合效率，但是甲醇浓度太高会影响蛋白质从凝胶中转移出来，尤其是高相对分子质量的蛋白质。半干转印仪和槽式的湿法转印仪都可以用来转印，但半干转印仪的转印效率不如湿法转印仪。

d) Western blotting 反应

转印完成后转印膜要进行封闭及与抗体反应。由于抗磷酸氨基酸抗体可以与牛奶中的某些成分结合，所以不能使用常用的的脱脂奶粉(Blotto)作为封闭液。磷酸化蛋白质的 Western blotting 反应可以使用牛血清白蛋白 BSA(组分 V)，或牛血清白蛋白与鸡卵白蛋白的混合物作为封闭液。一抗反应结束后使用针对第一抗体的辣根过氧化物酶偶联的第二抗体孵育，并用 ECL 检测系统检测磷酸化蛋白质。

e) 蛋白质的鉴定

为了鉴定检测出的磷酸化蛋白质，可以平行运行两块 2-D 胶。其中一块作为分析胶，另一块作为制备胶。将分析胶分离的细胞蛋白质电转印到 PVDF 膜上，进行免疫印迹反应，检测出磷酸化蛋白质，再染色显出其他全部蛋白质。制备胶分离的蛋白质直接染色后，与分析胶转膜后的谱图及免疫印迹后的 X 胶片进行对比，在制备胶上找出相应的磷酸化蛋白质进行进一步分析、鉴定。图 7-2 是这一方法策略的基本流程图。

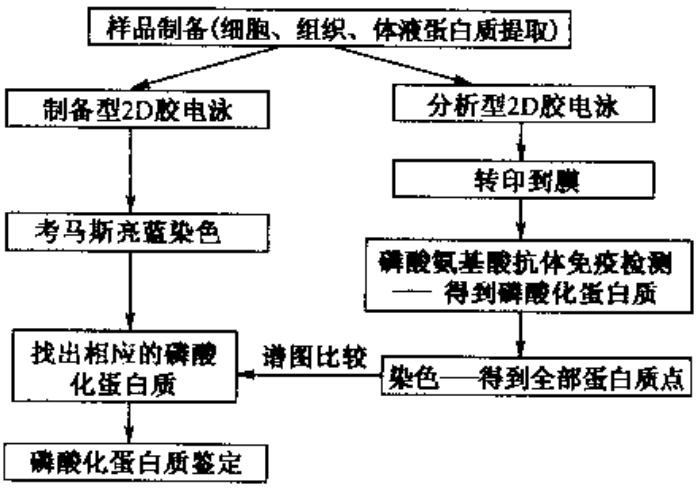


图 7-2　磷酸化蛋白质组的抗体免疫印迹分析策略

抗体免疫印迹检测法也可以与免疫沉淀技术相结合。即蛋白质在进行电泳分离之前先与固相化的抗磷酸氨基酸抗体进行免疫沉淀反应，免疫沉淀物洗脱后再进行电泳分离和 Western blotting 反应，这样在电泳之前可以先富集目的蛋白质，但这样也有可能丢失部分低丰度的目的蛋白质。

在目前有关蛋白质磷酸化修饰的蛋白质组研究工作中，用得较多的技术就是抗体免疫印迹检测分析方法。这些工作主要集中于研究在不同生长因子、细胞因子或其他因素的刺激下，细胞内蛋白质磷酸化状态的变化，进而揭示有关的信号转导途径或生理生化机理[5~11]。

下面举一个例子予以说明。Xavier F Csar 等[11]用巨噬细胞集落刺激因子(M-CSF)处理鼠造血细胞 FDC-P1 的野生型(FD/WT)和 CSF 受体上 807 位氨基酸突变型(Y807F)，通过研究细胞蛋白质的磷酸化状态受 Y807F 突变的影响，鉴别与巨噬细胞分化相关的磷酸化蛋白质。CSF 受体突变型 Y807F 是将 FD 细胞内 CSF 受体 CSF-1R 序列中 807 位的酪氨酸取代为苯丙氨酸，这种突变型细胞失去了在 CSF 的作用下分化为巨噬细胞的能力。实验中，FD/807 和 FD/WT 细胞在含 1% 血清的基质中抑制生长 18 h(未刺激)，然后用 CSF-1 处理 5min(刺激)，细胞蛋白质用二维凝胶电泳分离，转印到硝酸纤维素膜上，用抗磷酸酪氨酸抗体检测。通过比较未刺激的 FD/WT 和 CSF 刺激后的 FD/WT、FD/807 细胞酪氨酸磷酸化蛋白质，发现三种类型的磷酸化蛋白质：磷酸化状态与 CSF 受体 Y807F 突变无关的蛋白质，Y807F 突变使磷酸化程度增强的蛋白质及 Y807F 突变使磷酸化程度减弱的蛋白质。进一步的研究证实，在 Y807F 突变使磷酸化程度显著减弱的蛋白质中有 $p46/52^{Shc}$蛋白质，而且发现 $p46/52^{Shc}$可能在 CSF-1 诱导的巨噬细胞分化过程中发挥重要作用。

四、蛋白质磷酸化位点的分析

1. 磷酸肽的分离和富集

通过^{32}P 放射性标记或抗体免疫印迹方法检测出感兴趣的磷酸化蛋白质后，为了更清楚地了解蛋白质的磷酸化状态，例如磷酸化位点和磷酸化数目，需要对目标蛋白质进行进一步的分析研究。常用的分析方法是将蛋白质酶解成肽段，找到被磷酸化修饰的肽段，并对该肽段进行序列分析，确定发生磷酸化的氨基酸位点。一般来说，在分析蛋白质的磷酸化位点之前，要先鉴定该蛋白质，也就是要知道此蛋白质的序列。对于凝胶电泳分离的细胞或组织蛋白质，可以通过胶上原位酶切及肽质量指纹谱分析或串联质谱测序的方法鉴定，这种方法适合于微量样品的高通量鉴定。

蛋白质鉴定完成后从酶解肽混合物中找出磷酸化肽段并进行序列分析，在目前仍然是一个比较难以解决的问题。用经典的^{32}P 放射性标记方法时，一般通过

二维薄层色谱或反相液相色谱分离肽混合物，在分离过程中检测放射性比活度，收集有放射性活度的肽段进行序列分析。

虽然质谱技术的发展使得通过母离子扫描等串联质谱技术可以直接从肽混合物中选择性地分析磷酸化肽段并进行序列分析，但是由于蛋白质磷酸化的化学计量值较低，磷酸化肽段的信号丰度会比其相应未磷酸化肽段的丰度要低，而且在复杂的肽混合物中，质谱对磷酸肽的信号响应会被抑制。所以在分析前对磷酸肽进行分离或富集会使分析相对容易些。以下是一些用于分离富集磷酸肽的技术，分别予以介绍。

(1) 反相液相色谱(RP-HPLC)分离磷酸肽

经 RP-HPLC 分离，肽混合物成分可以被简化，磷酸肽可根据其疏水性被分离[3,12]。对于微量样品，最好使用毛细管液相色谱柱及极低的流速。对于^{32}P放射性标记的样品，分离收集的组分可经放射性比活计数检测含磷酸肽的组分。需要注意的是，磷酸肽的疏水性较弱，有可能某些高亲水性的磷酸肽洗脱时会直接流过柱子而不被检测到。RP-HPLC 系统可以与电喷雾(ESI)质谱在线连接，使用 HPLC-MS/MS 串联系统，结合串联质谱的各种扫描方式可以直接检测和鉴定肽混合物中的磷酸肽，尤其适用于没有被放射性标记的分析样品。

(2) IMAC 分离/富集磷酸肽

固相金属亲和色谱(immobilized metal affinity chromatography, IMAC)原来用于纯化含组氨酸的蛋白质，1986 年 Muszylnska G 等[13]发现固相化的铁螯合离子对磷蛋白和磷酸化氨基酸有高选择性的结合能力，现在 IMAC 也常被用来选择性富集磷酸肽[3,12,14-17]。IMAC 柱填料的基本构成是：色谱填料-交联剂-金属螯合剂，色谱填料常用琼脂糖凝胶，与金属螯合剂亚氨基二乙酸(iminodiacetic acid)或次氮基三乙酸(nitrilotriacetic acid, NTA)交联成为固定相，常用的金属离子 Fe^{3+}、Ga^{3+}和 Cu^{2+}等被螯合到固定相上。磷酸基团与固相化的金属离子有高亲和力，可被选择性地吸附在上面，用高 pH 溶液或磷酸盐洗脱后即为富集的磷酸肽。这一方法的局限性在于，有可能会丢失一些与 IMAC 柱结合能力较弱的磷酸肽，或某些有多个磷酸化位点的磷酸肽难以洗脱，另外那些富含酸性氨基酸链的非磷酸化肽段与固相金属离子也有结合能力，也有可能被富集。

(3) 磷酸基团亲和取代

将磷酸肽上的磷酸基团用另一种亲和配基取代，再用亲和提取的方法从混合物中分离富集磷酸肽，这是近两年出现的一种新技术。目前有 2 种亲和配基取代方法。一种方法[18]是使磷酸基团在碱性条件下发生 β 消除，然后由巯基乙醇取代磷酸基团的位置，最后在巯基上通过交联剂连接上生物素，生物素取代的磷酸肽由亲和素色谱柱分离，图 7-3 为示意图。这种方法只适合于磷酸化丝氨酸和磷酸化苏氨酸磷酸肽的取代。另一种方法[19]是通过化学反应在磷酸基团上连接一个半胱胺基团(cysteamine)，修饰的磷酸肽用固相化的碘乙酰凝胶亲和提取。这种方法

适合各种磷酸氨基酸的修饰。这两种亲和取代方法都需要进行化学反应，为了避免副反应的影响，都需要对蛋白质中的一些活性基团进行保护，所以整个反应体系还是比较复杂的。

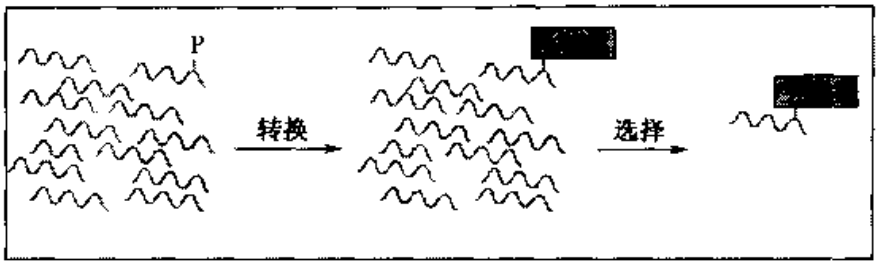


图 7-3　磷酸基团亲和取代[18]

2．磷酸肽的识别

对于没有进行^{32}P放射性标记的样品，要识别肽混合物中的磷酸肽，最好的方法是用质谱技术，或者质谱技术结合磷酸酶水解的方法。用串联质谱仪的母离子扫描、中性丢失扫描等检测方式，通过分析含磷酸基团肽段产生的特征离子，可以直接从肽混合物中找到磷酸化肽段。而用 MALDI-TOF-MS 结合磷酸酶的水解也可以找到磷酸肽。下面分别介绍在磷酸肽分析中应用到的各种质谱技术。

（1）MALDI-TOF-MS 结合磷酸酶水解分析法

这是蛋白质磷酸肽鉴定的常用手段。首先将磷酸化蛋白用蛋白水解酶水解为肽片段，再用磷酸(酯)酶处理肽段，使肽段脱磷酸。磷酸化氨基酸残基脱磷酸后，肽段失去了一个 HPO_3 分子，相应肽段的质量减少 80 Da，用质谱分析磷酸酶作用前后肽质量谱的差异，寻找质量数减少 80 Da 或 80 倍数的肽段。这样不仅可以用 PMF 的方法鉴定蛋白质，又可基本确定发生磷酸化的肽段及磷酸化的数目[20－32]。肽谱分析最常用的质谱仪是 MALDI-TOF- MS 质谱仪。图 7-4 是这一方法的简要示意图。

Martin R Larsen 等[20]设计了一种可以直接在 MALDI-MS 的样品靶上进行肽混合物磷酸酶处理的方法。具体过程是，先用 PMF 方法鉴定磷酸化蛋白质，然后将 MALDI-MS 样品靶上的基质样品结晶重新溶解，溶解液是 1.5μl 的 50mmol/L NH_4HCO_3 溶液，其中含有 0.05U/μl 的碱性磷酸酶。溶解后将样品靶放在一个密闭的塑料盒中，盒内放有湿润的织物以防止溶剂挥发干，在 37℃ 保温反应 30min，用 0.5μl 的 5% TFA 溶液终止反应，加入基质重新结晶并进行质谱分析。由于在肽混合物中的磷酸肽的质谱信号会被抑制，而且磷酸酶反应体系中含盐较多，所以在质谱分析前还有必要用微量脱盐柱脱盐。用这一方法分析溶液酶解的标准蛋白灵敏度在 20～100fmol，而对于电泳胶分离的蛋白质，要求胶上蛋白质的量至少要

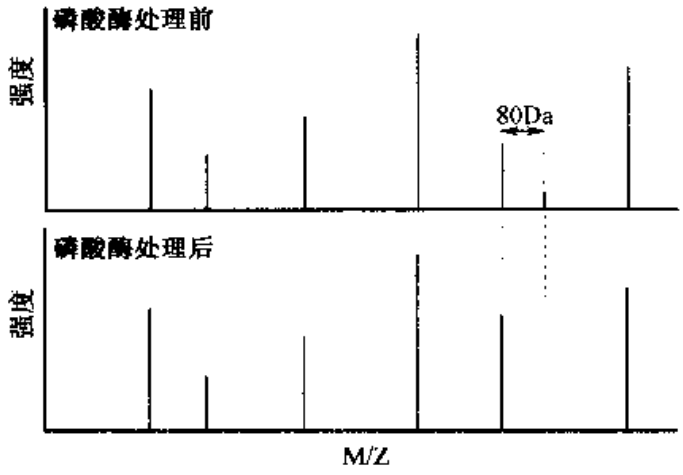


图 7-4　MALDI-TOF-MS 结合磷酸酶分析磷酸肽

增加 10～20 倍。

(2) PSD-MALDI-MS 分析

源后衰变是指发生在源内离子化之后的分子裂解，也就是 MALDI-TOF-MS 离子源内产生的离子在真空无电场飞行管道飞行过程中发生结构断裂，丢失一个中性分子后产生了碎片离子。由于是在无电场区飞行，所以碎片离子都具有与其前体离子相同的飞行速度，但由于质量小，所以其动能比前体离子的动能小，称为亚稳离子。由于前体离子和碎片离子具有相同的表观质荷比，飞行速度也一样，所以将同时到达线性检测器，在线性 TOF 分析器中，不能分辨碎片离子与前体离子。而碎片离子与前体离子存在动能差异，所以在具有反射器的 MALDI-TOF-MS 质谱仪上可以通过改变反射器电压分析到碎片离子。

磷酸化丝氨酸和磷酸化苏氨酸残基容易发生 β 消除反应，丢失 HPO_3 或 H_3PO_4 分子，产生质量数减少 80 Da 或 98 Da 的碎片离子(亚稳离子)，分别为 $[M+H-HPO_3]^+$ 和 $[M+H-H_3PO_4]^+$，通过源后衰变分析得到碎片离子峰的质量，寻找质量数减少 80 Da 或 98 Da 的肽段来判断磷酸肽[23]。虽然磷酸化酪氨酸残基不发生 β 消除反应，但是在 PSD 碎片离子谱图中也出现减少 80 Da 或 98 Da 的碎片离子[24]。

(3) 基于 MALDI-TOF-MS 的其他分析方法

Muller DR[25]和 Stensballe A[16]报道了用 MALDI-TOF-MS 观察去磷酸化后磷酸肽亚稳离子的分析方法。他们并没有用 PSD 的分析方式，而是直接用反射(reflector)方式收集肽谱，磷酸肽亚稳离子在肽混合物谱图中表现为一些峰较宽、分辨率很低的峰。基于亚稳离子峰低分辨率的特征，使其很容易与其他高分辨率的肽谱峰区分。由于质谱图中亚稳离子峰的表观质荷比并不代表其真正的质量，所以亚稳离子与其前体离子表观质量值之差并不是它们之间真正的质量差值。

Muller DR[25]的实验中观察到丢失 98 Da 的亚稳离子峰与其前体离子的表观质量数之差约为 94 Da,而 Stensballe A[16]的实验观察到的差值为 86 Da,可能亚稳离子的表观质量与质谱仪的加速电压和反射电压值有关。图 7-5 是亚稳离子与正常分辨离子峰的比较。

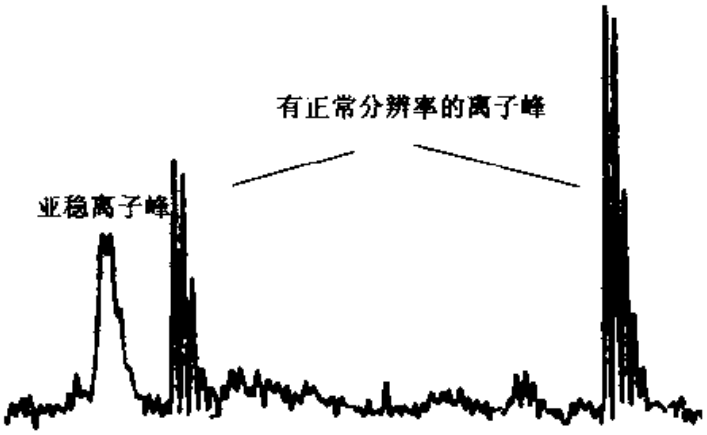


图 7-5　亚稳离子与正常分辨率离子峰比较[29]

John M 等[26]报道在磷酸肽与非磷酸肽混合物与基质共结晶时,在基质溶液中加入铵盐,如柠檬酸铵,用 MALDI-TOF-MS 正离子检测模式分析,发现磷酸肽的响应信号明显增强,用不同的基质如 4-HCCA 和 DHB 得到的结果相近。可通过比较加入铵盐前后肽谱图的差异来确定磷酸肽。

Ma Yuliang 等[27]报道利用 MALDI-TOF MS 正负离子检测模式对磷酸肽响应的差别来确定磷酸肽。在负离子检测模式下,磷酸肽的响应信号要比正离子检测模式下的响应信号相对要强。通过比较正负离子检测模式得到的肽谱信号相对强度的变化,能够发现磷酸化的肽段,再对找到的目标肽段用串联质谱测序进行验证。作者用这一方法明确指认出 β 酪蛋白酶解产物中的两个磷酸肽,对溶液酶解的 β 酪蛋白样品,检测限可达到 18 fmol。作者用这一方法结合串联质谱测序法鉴定了 PKR 样内质网激酶(PERK)的多个自磷酸化位点。

前文所讲述的固相金属离子亲和色谱法也是选择性检测磷酸肽的方法。固相化金属螯合色谱填料有一个好处是可以直接用 MALDI-MS 分析[17],即直接用此 IMAC 填料与肽混合物共孵育,使磷酸肽与色谱填料结合,离心后填料与磷酸肽共同沉淀在底部,弃上清,将沉淀清洗几遍后,可直接与 MALDI-MS 的基质溶液混合并进行质谱分析。或者使用 IMAC 柱将磷酸肽富集洗脱后质谱分析,通过比较 I-MAC 处理前后肽谱的变化找到磷酸肽。

IMAC 柱可以自己装填也可以使用装填好的商品化柱子,Millipore 公司已经有了用于富集微量磷酸肽样品的 ZipTip 小柱,IMAC 填料装填在 10ml 的 pipette 枪头。图 7-6 是一个使用 ZipTip 金属离子亲和小柱提取磷酸肽的例子。磷酸化蛋

白质β酪蛋白经 SDS-PAGE 凝胶电泳分离后进行胶上酶切，肽混合物用 Fe^{3+} 螯合的 ZipTip 金属离子亲和小柱提取磷酸肽，在含 0.1% 醋酸的酸性溶液中使磷酸肽吸附在 ZipTip 柱上，用酸性溶液洗脱其他未结合的肽段，最后用 0.3mol/L 的氨水洗脱磷酸肽，用 MALDI-TOF-MS 分析 ZipTip 处理前后的肽谱。可以看到，经 ZipTip 亲和提取后肽谱被明显简化，而且原来肽混合物中的质荷比为 m/z 2061.9 的一个小峰成为基峰，这一肽段就是β酪蛋白酶解产物肽段中的第 6 段，蛋白质氨基酸序列中的 33 到 48 位，序列为 FQSEEQQQTEDELQDK，该肽段质荷比的理论值为 m/z 1981.9，但是在其中的第 35 位丝氨酸残基上有一个磷酸基团，使该肽段的质量数增加了 80 而成为质荷比为 m/z 2061.9 的峰，并被 Fe^{3+} 螯合的 ZipTip 亲和小柱特异性吸附。

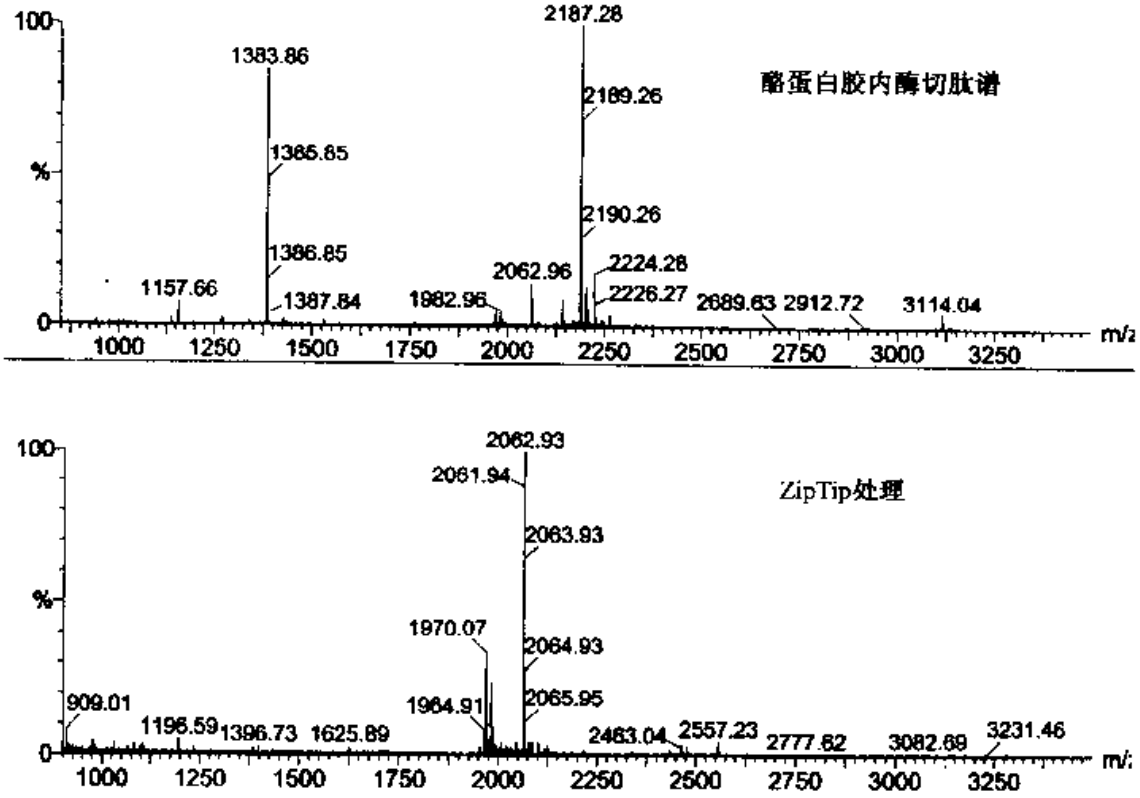


图 7-6　Fe^{3+} 螯合 ZipTip 亲和柱提取磷酸肽

(4) 中性丢失扫描

具有两级质量分析器和碰撞诱导裂解室的串联质谱仪可以用中性丢失扫描方式(neutral loss scan mode)检测具有某一特征结构特点的一类离子。在三级四极串联质谱仪中采用此种扫描模式时，第一个(Q1)和第三个(Q3)四极杆质量分析器同步扫描，但扫描范围不一样，保持一个特定的电压差值，这个电压差所代表的质荷比 m/z 值代表某一中性分子的质量。第二个四极杆(Q2)是碰撞室，将 Q1 输送过

来的离子碰撞诱导裂解，再将碎片离子送入 Q3。这样，经 Q1 输送的混合物样品离子中，只有在 Q2 中诱导裂解后能丢失这一特定中性分子的离子才会被 Q3 传输到检测器。分析磷酸肽时，如果从带一个正电荷的$[M+H]^+$肽混合物中寻找丢失中性磷酸分子 H_3PO_4 的磷酸肽，则 Q1 和 Q3 的扫描电压差所代表的质荷比就应该是 m/z 98，如果从带两个正电荷的$[M+2H]^{2+}$肽混合物中寻找丢失中性磷酸分子 H_3PO_4 的磷酸肽，则 Q1 和 Q3 的扫描电压差所代表的质荷比就应该是 m/z 49[28,29]。此方法的优点是以正离子模式进行扫描分析，找到磷酸肽后可以直接通过碰撞诱导解离分析磷酸肽的序列及磷酸化位点。此方法的局限在于只能分析磷酸化丝氨酸和磷酸化苏氨酸，因为只有磷酸丝氨酸和磷酸化苏氨酸会发生 β 消除而产生 H_3PO_4 的中性丢失。在一次扫描分析中只能检测带一种电荷的离子，对于带多种电荷的混合物离子要进行多次扫描分析，另外分析未知混合物时有可能出现假阳性结果。

(5) 前体离子扫描

前体离子扫描又称为母离子扫描(precursor/parent ion scanning)，在三级四极串联质谱仪中采用此种扫描模式时，连续扫描第一个四极质量分析器 Q1，使各种质荷比的离子依次通过 Q1 分析器，Q2 为碰撞诱导解离室，将 Q3 分析器的电压和频率设定为只能传输某一特定质荷比的离子，即特定质荷比的子离子。与中性丢失扫描不同的是，前体离子扫描模式直接检测母离子在碰撞诱导解离过程中失去的碎片离子。检测器得到碎片离子的信号时，已知该碎片离子的前体离子通过 Q1 时的电压，根据 Q1 电压值确定该前体离子的质量。因此，前体离子扫描模式被用来检测肽混合物中某些含有特定结构特征的肽段。

分析磷酸肽时，在 ESI 负离子模式下对 Q1 进行连续扫描，设定 Q3 中仅能通过的子离子质荷比为 m/z 79，即磷酸肽丢失的 PO_3^- 基团，这样得到的质谱图仅显示会丢失 m/z 79 的离子的谱峰，也即会丢失磷酸根离子的磷酸肽谱峰。因为多肽产生的各种碎片离子中，质量数在 79Da 附近的几乎没有，所以用这一质量数进行母离子扫描的特异性是很高的。但是此种扫描方式存在测序的问题，因为在负离子检测方式下进行序列分析的结果不好，所以检测到磷酸肽后，要立即切换为正离子模式进行测序。而通常在 ESI 负离子检测时，为了提高信噪比，样品都溶解在碱性溶液中，所以为了在正离子检测模式进行序列分析，需要更换样品溶液体系，或者将样品分成两部分，分别用不同的溶剂溶解[14]。如果连接了 HPLC 分离系统，在酸性的液相洗脱溶液中负离子检测模式的离子信噪比会很低，影响了分析的灵敏度。

酪氨酸磷酸肽会产生质荷比为 m/z 216.043 的特征离子，这是磷酸酪氨酸的亚铵离子 $C_8H_{10}NO_4P$[immonium ion, Im(pY)]，选这一离子作为子离子进行母离子扫描，可以在正离子检测模式下选择性检测酪氨酸磷酸肽，然后直接进行序列分析。但是由于各种氨基酸残基产生的碎片离子质量数在其附近的不少，而四极杆

质量分析器的分辨率又较低，如果用三级四极质谱作母离子扫描，选择性会很低。

四极杆-飞行时间串联质谱(Q-TOF)并不能实现真正意义上的母离子扫描，但是在扫描四极杆的同时用飞行时间检测器收集所有的碎片离子谱，同样可以将碎片离子的质量数与四极杆的质荷比位置，即母离子的质荷比对应起来。由于飞行时间检测器的高灵敏度和高分辨率，使 Q-TOF 质谱仪可以选择高精度的碎片离子质量数用于母离子扫描实验。但是，Q-TOF 串联质谱仪作母离子扫描的灵敏度却非常低，这主要是因为仪器对小分子碎片离子从碰撞室到 TOF 分析器的离子传输速率太低，只有 5%[30]。解决这一问题的主要设计方案是 Q2 pulsing，就是使子离子在进入 TOF 分析器之前在 CID 碰撞室内被俘获(trap)很短的一段时间，离子的释放与 TOF 分析器正交加速电场的脉冲电压同步，这样可以使小分子碎片离子的传输速率达到 95% 以上。在具有 Q2 pulsing 功能的仪器上(如 API QStar Pulsar Quadrupole Time of Flight with a LINAC collision cell)可以成功地进行母离子扫描分析[31]。

Steen H[32]对三级四极质谱仪与 Qq pulsing-TOF 质谱仪的母离子扫描功能进行了比较，结果表明二者均可以进行 m/z 79 的负离子模式母离子扫描，而 m/z 216.043 的正离子模式母离子扫描只适用于 Qq pulsing-TOF 质谱仪。他们还用 m/z 216.043 母离子扫描方法对 EGF 受体信号通路中的酪氨酸磷酸化蛋白质进行了鉴定和磷酸化位点分析[10]。正离子模式母离子扫描的最大好处就是可以直接对磷酸肽进行序列和磷酸化位点分析，甚至不必先知道蛋白质的序列。

3. 磷酸化氨基酸位点的确定

找到磷酸肽后，要确定其磷酸化位点，需要用串联质谱产物离子扫描模式(product ion scanning)对磷酸肽进行序列分析[3,10,22]。将选定的磷酸肽在碰撞诱导解离(CID)室打碎，检测其产生的全部碎片离子，根据碎片离子质量数推断肽段序列和磷酸化位点。磷酸肽在 CID 室碰撞诱导解离，易产生丢失 80 Da(HPO_3)和 98 Da(H_3PO_4)的子离子，此离子峰的丰度很高，往往在整个碎片离子谱中占据主导地位。一般较短的磷酸肽比较长的磷酸肽更易于丢失磷酸，可能是由于在相同碰撞能量下，短肽的化学键上分布了更多的能量。磷酸肽链骨架会断裂成为 y、b 等碎片离子系列，肽链骨架断裂产生的离子质量数有利于对肽链进行序列分析。带有磷酸基团的碎片离子也会产生丢失 98 Da(H_3PO_4)的峰，这个质量数会与非磷酸化碎片离子的失 H_2O(18Da)产物的质量数混淆，有时会给 CID 谱的分析和解析带来困难。图 7-7 是一个磷酸肽的串联质谱图例[14]。

用 HPLC 与 CE 连接的串联质谱会简化肽混合物的质谱峰，但是因为肽段洗脱时间有限，有时得不到好的 MS/MS 谱图，在有肽段共洗脱时，没有时间分析每一个肽段的序列，因此，在线(on line)分离-MS/MS 联用方法并不总能鉴定磷酸化位点。

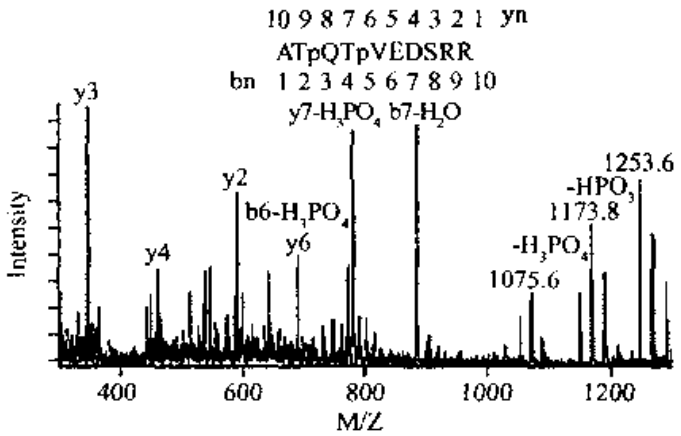


图 7-7　串联质谱磷酸肽测序图例[14]

五、蛋白质磷酸化的定量分析

蛋白质磷酸化的定量问题，即磷酸化蛋白质与其相应非磷酸化蛋白质的比例，也是蛋白质组研究非常关心的问题。早期研究磷酸化蛋白质定量是将磷酸肽与非磷酸肽分离，用 HPLC 分析磷酸化氨基酸，用峰面积定量比较磷酸氨基酸与非磷酸氨基酸的比例，计算磷酸肽的化学计量值。这种方法一般需要将磷酸肽用^{32}P放射性同位素标记，而且需要的磷酸化蛋白的量很大。要从蛋白质组的水平研究磷酸化蛋白质的定量，目前还没有实用、有效的方法。用稳定同位素标记的方法定量蛋白质表达，是近两年来出现的新技术，虽然方法独特，但还只停留在对标准蛋白质及简单蛋白混合物的分析水平，要应用在复杂的细胞蛋白质混合物研究上目前还很困难。下面分别介绍两种稳定同位素标记磷酸化蛋白质定量的方法。

(1) ^{15}N 稳定同位素标记法[33]：将两种细胞在不同的培养基中分别培养，一种为正常培养基，另一种培养基的 N 源由^{15}N提供，混合两种培养物，提取细胞蛋白。分离目标蛋白质(磷酸化蛋白)后，酶解，提取肽段，用质谱分析。因为两种细胞蛋白分别掺入^{14}N和^{15}N，所以在质谱图中，每个水解肽段表现为一对峰(见图 7-8)。$^{14}N/^{15}N$的同位素丰度比可以体现出两种细胞来源蛋白质表达量的相对水平，如果两种细胞表达完全一致，则所有肽段$^{14}N/^{15}N$丰度比将在一个平均值范围以内，而如果两种来源蛋白质的磷酸化程度不同，其含磷肽段与相应未磷酸化肽段的$^{14}N/^{15}N$丰度比就会与平均值不同，根据其比值来进行磷酸化程度的相对定量。这一方法需要在标记培养基中培养细胞，有可能培养基的差异本身就造成蛋白质表达量的变化，而且目标蛋白质需要纯化、分离，所以还不能用于大规模的蛋白质定量分析。

(2) 磷酸肽同位素亲和标记法[34]：使磷酸肽或磷酸化蛋白在碱性环境下发生

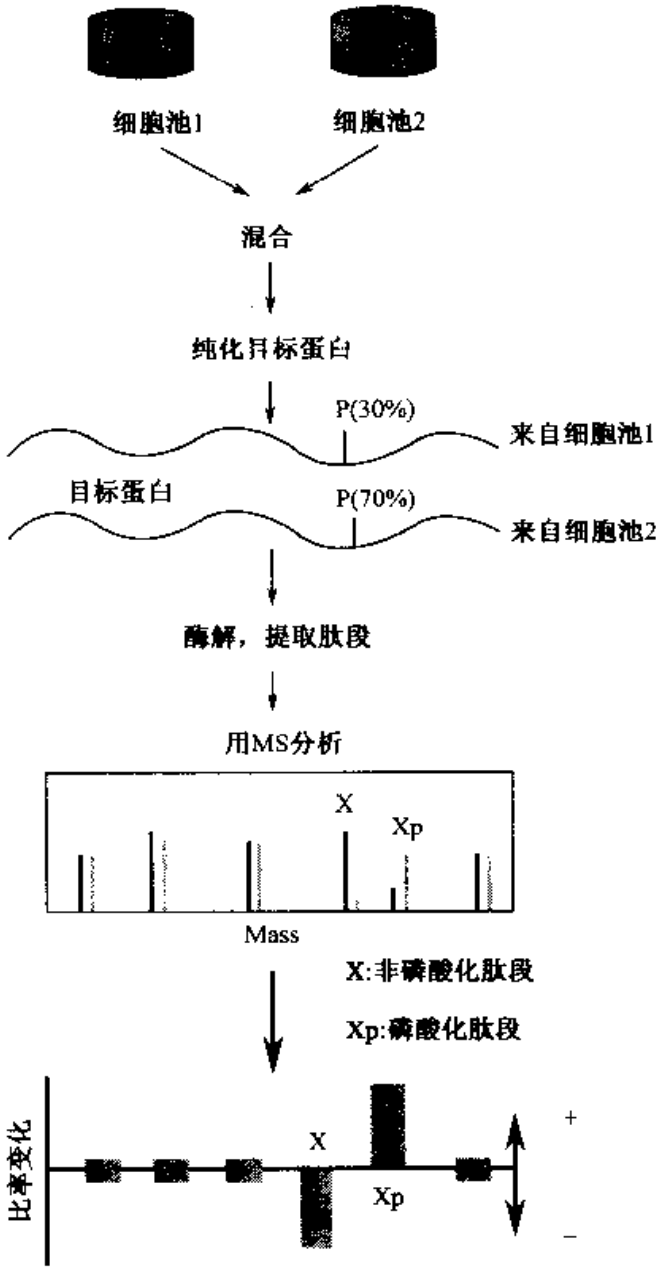


图 7-8 ^{15}N 稳定同位素标记法定量磷酸化蛋白质[33]

β消除，同时用亲核试剂攻击形成的双键并发生加成反应。对两种不同来源的磷酸肽或磷酸化蛋白使用不同的亲核试剂，区别在于，其中一种试剂上的氢原子由氘代替。再通过化学反应在加成的亲核试剂上连接一个亲和标签，例如生物素(biotin)。将两种细胞来源的磷酸肽或磷酸化蛋白混合，酶解后提取含亲和标签的肽段并用质谱分析(见图 7-9)。由于掺入了不同的亲核试剂，在质谱图中，磷酸肽表现为一对峰，同样根据其强度比进行定量。此方法的化学反应太多，而且需要纯化的步骤也多，目前仅能用于标准蛋白的定量分析。

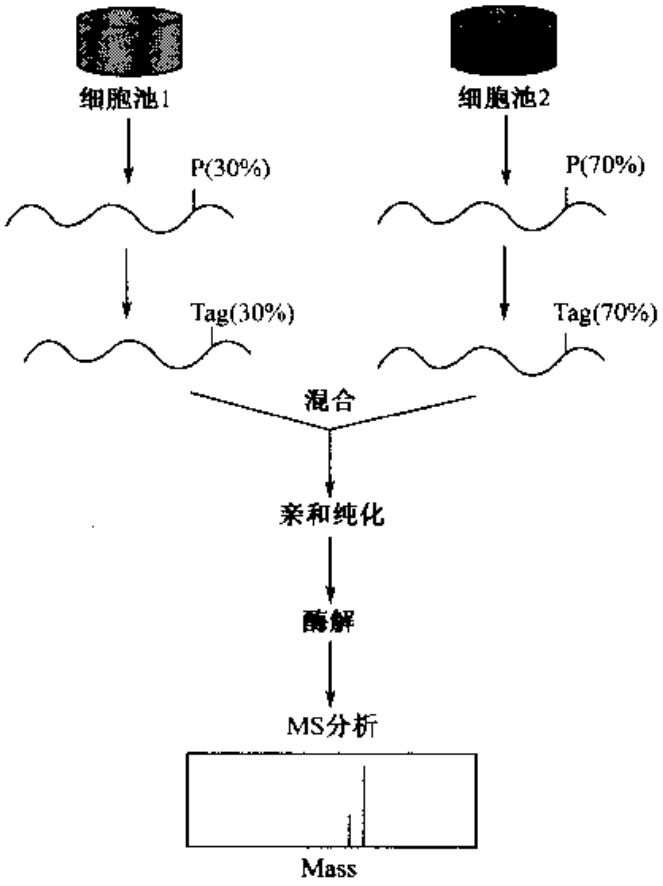


图 7-9　磷酸肽同位素亲和标记法定量磷酸化蛋白质[12]

六、小结

目前，现有的磷酸化蛋白质研究手段还不能满足研究的需要，尤其是对微量磷酸化蛋白的富集、鉴定、磷酸化的定量，还需要更特异地富集磷酸化蛋白和磷酸肽的方法。磷酸化修饰是蛋白质组研究中一个非常重要的内容，蛋白质磷酸化与其功能密切相关，所以磷酸化蛋白质组研究都应在一定的功能环境下研究磷酸化蛋白，例如在某一信号转导通路内或在某种外界刺激下细胞蛋白质磷酸化的变化。除了研究细胞总蛋白质外，还应该侧重于细胞某些特定部位的蛋白质磷酸化现象，例如与信号转导有关的许多细胞因子受体都为跨膜蛋白，所以膜蛋白应该成为一个主要内容研究。细胞核内的酪氨酸磷酸化现象及其功能网络目前还不是很明确，用蛋白质组技术体系来研究这一问题也许会取得一些有意义的结果[35]。

第二节　糖基化蛋白质的鉴定

糖基化蛋白质简称糖蛋白，广泛存在于动物、植物、微生物及病毒中。近20年来的研究发现，糖蛋白在各种生命现象中起着重要的作用，如：细胞识别、信号传导、免疫与应答、分化与反分化等等，都与存在于糖蛋白中的糖链的结构与作用有关。自19世纪末人们开始提出糖蛋白的概念并加以研究以来，先后明确了其糖基结构、成分、部分生物学功能，以及认识了糖蛋白中的糖苷键类型。到目前为止，随着生物工程技术的发展，也出现了多种颇具疗效的由真核细胞表达的糖蛋白类药物，如干扰素(IFN)、促红细胞生成素(EPO)等。由于糖蛋白本身的微不均一性，使得其结构、位点的差异直接影响到疗效。虽然对重组糖蛋白的质量或结构均一性的控制在某些方面是可行的，如：宿主细胞的选择、糖加工过程中的基因调控以及生物工程进程中参数(如培养时间)的控制等，但这些手段都不足以保证产生单一的糖蛋白产品。因此，无论是自然的还是重组的糖蛋白，其结构鉴定就成为了一个十分重要的课题。

长期以来人们沿用的分析糖蛋白的方法是将糖蛋白中的糖链及蛋白分开来研究，即利用化学方法如肼解法、氧化还原法或衍生物法先将糖链释放，然后用层析法，如Sephadex G-25或G-50进行凝胶层析分离，收集含糖组分，或用SDS-PAGE、HPLC等传统的分离分析方法，粗略地观察糖蛋白的分子质量。应用化学方法降解、还原多糖，从而研究糖链结构的方法虽具有普遍性，但反应周期长(几小时到几天不等)，反应条件难以控制，对糖蛋白的需要量非常大，而且并不是所有的糖蛋白都能适应这种高温、碱性的实验条件而不被降解。

随着现代分析技术的发展，许多分析手段都能够详细地分析糖蛋白及糖链的结构。在众多分析方法中，生物质谱法由于它的操作简便并具有极高的灵敏度和准确度，迅速成为了生命科学研究领域的新工具。20世纪80年代初期，随着许多软电离方式的出现，如FAB-MS(快原子轰击)、ESI-MS(电喷雾质谱)和MALDI-TOF-MS(基质辅助激光解吸附飞行时间质谱)等，使质谱法摆脱了只能分析小分子化合物的局限，延伸向生物大分子聚合物的测定，它能够快速、准确地提供蛋白质、核酸、聚合物等生物大分子的相对分子质量、序列、一级结构信息以及结构转换、修饰等方面的信息。同样，对于糖蛋白的分析，质谱法可以摆脱传统的分离分析方式，直接分析糖蛋白的相对分子质量、糖基化位点、氨基酸序列以及糖链结构等等，结合Internet网上数据库检索系统，为生命科学的研究提供高质量的结构信息。

蛋白质组学是一个以蛋白质为基础的整体生物学方法，在绝大多数情况下，一个成熟蛋白的生物学活性是被翻译后修饰的，而这些修饰又是高度动态的。在疾病的最初阶段，糖蛋白通常在结构与功能上显示出变化，而这些变化涉及到与蛋白

相连的多糖结构，人们已经开始讨论在人类疾病过程中糖基转移酶缺失情况下这些糖蛋白交互作用的机理[36]，影响并导致疾病进程的糖基化变化与真正引起疾病的变化二者之间的差异、可以充当干涉治疗的位点，也是人们最感兴趣的。

一、糖蛋白的结构特征

糖蛋白是一类由糖类和蛋白质以共价键连接而成的结合蛋白，修饰范围从蛋白质上连接单糖（比如胶原糖蛋白及核糖蛋白），到密集、复杂、有侧链结构的多糖（如黏液素），以及具有复杂的氨基葡聚糖侧链的蛋白多糖。多糖与蛋白质相连有两种最基本的形式：一类是β-构型的 *N*-乙酰氨基葡萄糖（β-GlcNAc）与天冬酰胺（Asn）的氨基形成的 N-糖苷键（图 7-10），其中 Asn 经常处于 Asn—X—Thr/Ser（X 是除 Pro 以外的任何氨基酸）的序列子中，我们称其为天冬酰胺序列子，只有处于该序列子中的 Asn 才会发生糖基化。

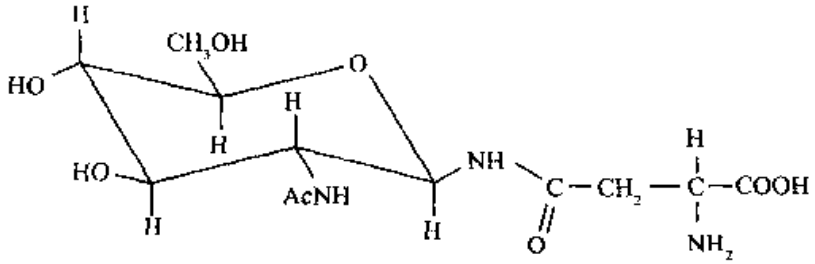


图 7-10　N-连接糖肽链[37]

N-糖链结构通常发生于内质网及高尔基体上，其内侧通常包含一个五糖核心，即三甘露糖五糖核心 $GlcNAc_2Man_9$（图 7-11），在此基础上，由糖苷酶及糖基转移酶将其进行多样化整理和修饰，几乎所有的 N-连接多糖内侧都具有五糖核心结构，它与天冬酰胺序列子中的天冬酰胺残基相连接。按照单糖的位置可将 N-连接结构分为高甘露糖型、混合型或复杂型多糖。不同的糖链可连于同一位点（微不均一性），同时在同一蛋白质上也可有不同的位点连接糖链（显著不均一性）。

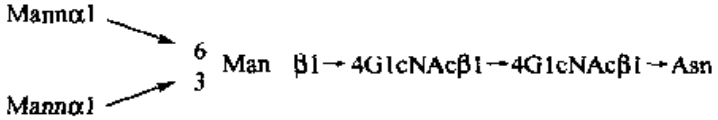


图 7-11　N-糖链五糖核心区[37]

另一类是 O-糖苷键，它最主要是由 GalNAc 的异头碳与 Ser 或 Thr 中的羟基缩水形成（图 7-12），这一类型的糖苷键占 O-糖苷键的主体，最初是从颌下腺、消

化道、呼吸道等黏液糖蛋白中发现，故又称为黏蛋白型糖苷键。但近来在人免疫球蛋白、人绒毛膜促性腺素、胎球蛋白等非黏液型糖蛋白中也相继发现有此种糖苷键。除 GalNAc 外，其他一些糖，如 GlcNAc、Gal、Xyl、Ara、Man 也可与 Ser 或 Thr 相接，这些类型的 O-糖苷键存在于动物细胞浆或核浆、玉米根蛋白聚糖、植物细胞壁、酵母细胞壁及地蚕表皮胶原中等等。另外 L-Ara 的异头碳与羟脯氨酸的羟基缩水，或半乳糖的异头碳与 5-羟赖氨酸的羟基缩水都可形成 O-糖苷键，它们的特点是对碱稳定。O-连接糖链结构的共同点是由一种或少数几种单糖与某些含羟基氨基酸连接，与 N-连接糖链结构不同，后者并无一个简单的规律认为含羟基残基可以作为连接位点[37]。

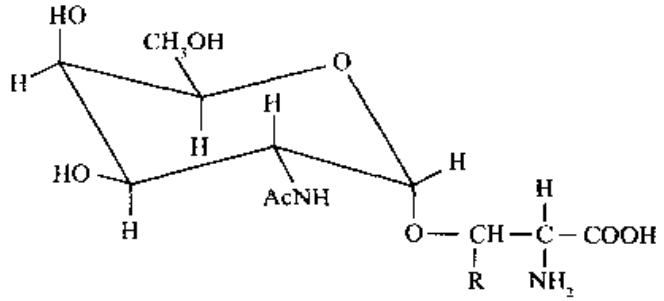


图 7-12　O-连接糖肽链[37]

还有一种特殊的糖基磷脂酰肌醇(GPI)固定结构，这种结构具有将细胞膜的疏水性磷酸双分子层固定于不同蛋白质的功能[38]。它们具有相同的 Man_3GlcNH_2 糖核，是细胞膜内磷脂与蛋白质之间的桥梁，这一翻译后修饰行为发生在内质网中信号肽 C 端切割之后。在同样一个蛋白质上可以发现不同的修饰方式共存，这是其宏观不均一性；同时，同一修饰点上还存在着微不均一性，这是因为肽的连接都是氨基酸的 α-氨基和 α-羧基连接成肽键，一个氨基酸残基只能在氨基侧或羧基侧形成一个肽键，一般不会形成分支肽链，但寡糖中两个糖基的互相连接可以有 1→2、1→3、1→4、1→6 等不同方式，一个糖残基和相邻残基有时可形成 4 个糖苷键，从而使糖链分支，而且糖基还有 α、β 异头构型，更造成了连接键的复杂性。

二、生物质谱技术鉴定糖蛋白

在众多软电离技术中，FAB 是最先应用于生物分子测定的，对于小分子含糖化合物如糖肽可直接测定，但由于这种仪器在分子质量范围方面的局限性、少量碎片离子的产生以及仪器价格的昂贵使它的应用受到了限制，而 ESI-MS 及 MALDI-TOF-MS 以其操作简单、耐用、灵敏度好等优势已被越来越广泛地应用。

糖含量、糖苷键类型及糖基化位点是糖蛋白一级结构的重要指标，应用质谱技术分析糖蛋白的思路如图 7-13。在糖蛋白中，糖链一般处于蛋白质高级结构的外侧，糖苷内切酶切断糖链与蛋白链间连接的糖苷键，因此，用糖苷内切酶将糖链切除，将反应前后的质谱图进行比较，就能直接表述糖链的平均质量，而糖蛋白的平均糖含量可由糖链的平均质量占糖蛋白平均相对分子质量的百分比来表示。糖基化位点是糖蛋白结构中极其重要的特征之一，在糖蛋白中，可以有一个或多个糖基化位点，采用蛋白酶酶解与糖苷内切酶酶解相结合的方法，先用蛋白酶将糖蛋白酶解成含有和不含有糖链的肽片段，获得糖蛋白的肽质量指纹谱，再用糖苷内切酶将糖链切除，其肽质量指纹谱将发生变化，原含糖肽段的质量由于糖链的丢失在质谱图上发生了位移，通过网上数据库检索可以得到位移肽片段的氨基酸序列，而在这个序列中必含有糖基化位点，由此我们可以确定糖蛋白分子中的糖基化位点。应用电喷雾串联质谱技术，对肽质量指纹谱中的肽片段进行分析，可以得到蛋白某中间片段的序列，对已知蛋白，依靠此序列，判断其归属，从而对糖蛋白的氨基酸序列加以证实。凝集素是一类糖结合蛋白，能专一地识别某一特定结构的单糖或寡糖中特定的糖基序列而与之结合，它不但可用于寡糖或糖复合物的分离纯化，也可用于糖链的结构分析，因此，用凝集素直接提取糖肽段，可以同时起到富集、浓缩以及分离的多重效果。寡糖链的结构在糖蛋白的生物学活性中起着极其重要的作用，随着串联质谱的发展，结合计算机软件以及 Internet 网上数据库检索系统，已经能够应用质谱的方法直接分析糖蛋白中的寡糖链[39,40]。

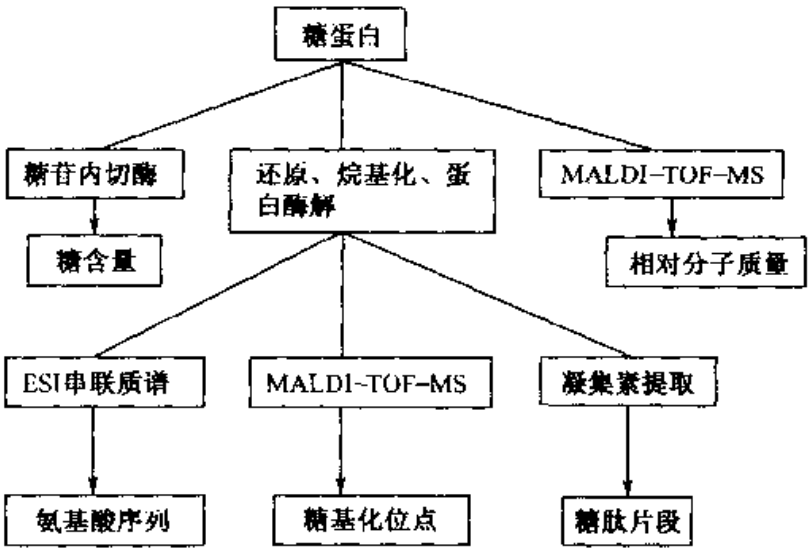


图 7-13　糖蛋白结构分析示意图

1. MALDI-TOF-MS 的应用

基质辅助激光解吸附飞行时间质谱仪(MALDI-TOF-MS)以具有强紫外吸收的小分子有机酸作为基质，与蛋白样品以 5000∶1 的比例混合加在不锈钢靶上形成共结晶薄膜，基质吸收紫外激光(N_2 激光源，波长 337nm)的能量，形成激发态，导致蛋白的电离和气化，在电场加速下进入真空飞行管道，从而到达检测器，绝大多数离子可以带 +1 到 +3 个电荷，谱图一般为多次扫描的累加。这种仪器的优势在于简便、直观，大分子混合物不经分离可以直接进行测定，基本不产生碎片峰，同时，它的灵敏度之高是其他类型质谱仪所不能达到的。

(1)基质的选择

在 MALDI-TOF-MS 分析中，基质起着相当重要的作用。到目前为止，虽然对基质在 MALDI-TOF-MS 中的作用机理并不十分清楚，但却发现了许多效果颇佳的基质，如：α-腈基-4-羟基肉桂酸、芥子酸、2,5-二羟基苯甲酸、2,4,6-三羟基苯乙酮、烟酸、7-羟基-4-甲基香豆素等，基本上都是水溶性较好的小分子有机酸。不同的基质对不同类的物质响应不同，如：肉桂酸、芥子酸类基质适合于蛋白质，2,5-二羟基苯甲酸适合于聚合物，而 2,4,6-三羟基苯乙酮适合于核酸类物质。我们选择了几种适合于与糖蛋白结构性质相类似的蛋白质的基质，用它们测定核糖核酸酶 B 这一糖蛋白标准品，比较它们对糖蛋白的响应程度(图 7-14)。从图中

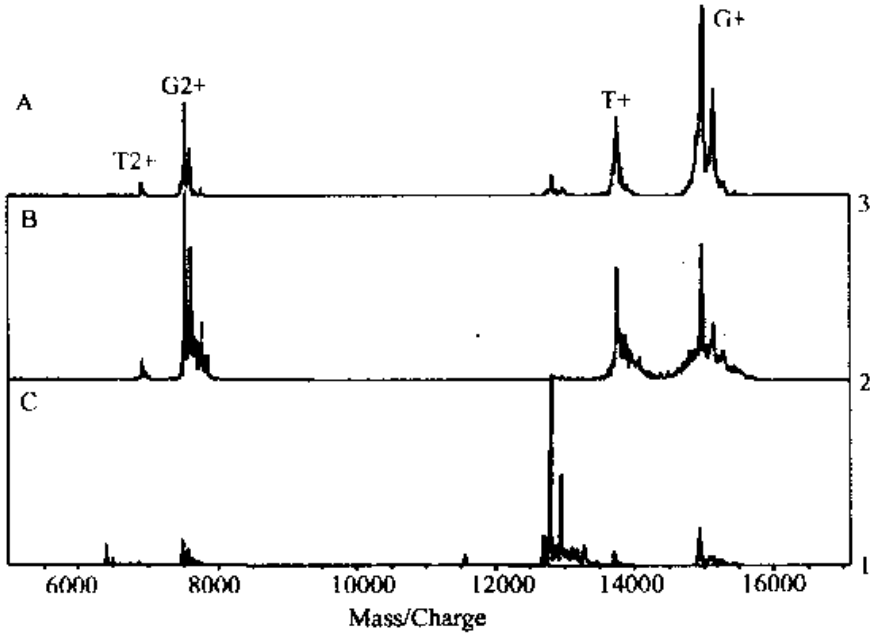


图 7-14　不同基质作用于核糖核酸酶 B 的 MALDI-TOF-MS 图

A.α-腈基-4-羟基肉桂酸。B.2,5-二羟基苯甲酸。C. 芥子酸。G^+、G^{2+} 为糖蛋白分子离子峰及双电荷峰，T^+、T^{2+} 为少量无糖蛋白质分子离子峰。

可以看出，芥子酸对糖蛋白的响应极弱；2,5-二羟基苯甲酸对糖蛋白样品中少量无糖蛋白的响应较好，同时削弱了糖蛋白的丰度，也不适合于糖蛋白；只有 α-腈基-4-羟基肉桂酸对于糖蛋白(G^+)及其少量无糖蛋白(T^+)的响应丰度符合于实际情况，它对糖蛋白的响应较稳定，并且重复性好。而其他样品分析结果也显示 α-腈基-4-羟基肉桂酸用于测定糖蛋白效果相对较好。

(2)糖蛋白的平均分子质量及糖含量测定

与 SDS-PAGE 或 HPLC 等传统的测定相对分子质量的方法不同，质谱法以质荷比的方式将相对分子质量不同的物质从小到大排列开来，在质谱图上，横坐标为分子离子的质荷比，纵坐标为相对丰度，从各个质谱峰的位置，可以一目了然地得知其相对分子质量。同时，在各种软电离方式中，多电荷峰的出现还可以辅助证明分子离子峰的存在。因此，在测定相对分子质量方面，质谱法比其他传统方法更加灵敏、准确、直观。

与普通蛋白不同，糖蛋白分子结构具有不均一性，包括糖链的微不均一性及糖基化位点的不均一性、多羟基结构的糖链使其质子化能力降低、糖链末端唾液酸的干扰等，这些特点给糖蛋白的质谱测定带来了困难，但应用性能高、灵敏度好的 MALDI-TOF-MS 也可以直接测定分子质量不大、糖链不十分复杂的糖蛋白如核糖核酸酶 B、α1 酸性糖蛋白等[41]，而对于分子质量大(大于 50kDa)、糖链复杂的糖蛋白如胎球蛋白，就很难直接测定其分子质量。但依靠适当选取基质以及结合相应的制靶技术，可改善对某些糖蛋白的质谱测定，如利用靶上清洁基质法[42]，可以提高测定重组糖蛋白的灵敏度；同时，唾液酸的存在通常会干扰糖蛋白的测定，可以预先将唾液酸用唾液酸酶切掉，或将唾液酸甲基酯化[43]，再进行测定，从而提高测定灵敏度，同时也可增加糖链的稳定性。

核糖核酸酶 B 由 124 个氨基酸组成，在 34 位 Asn 处连有一个高甘露糖型 *N*-糖链，我们在 MALDI-TOF-MS 上测定了其相对分子质量(图 7-15)，由于糖链的微不均一性，与普通蛋白质及核酸不同，其分子离子峰在 MALDI-TOF-MS 质谱图上表现为一簇峰 G^+，各峰之间约相差一个糖基，同时还出现了双电荷峰 G^{2+}，其中 T^+ 及 T^{2+} 为样品中含少量不带糖链的蛋白峰及其双电荷峰。根据 G^+ 五重峰的测定值，可以直接计算出糖蛋白的平均相对分子质量($X = \sum X_i/n$)。正是由于糖蛋白结构的这种微不均一性，使得其分子离子峰变宽，检测灵敏度降低，糖链分子质量越大，峰越宽，灵敏度越差，所以，只有那些糖链较短(比如小于 5kDa)，蛋白的质量不太大(比如小于 150kDa)的糖蛋白在性能较好的仪器上才能测定其平均分子质量。

糖苷内切酶切断糖链与蛋白之间的糖苷键，用这种酶切方法可以直接测定糖蛋白的糖含量。有些糖蛋白样品中含有不带糖链的蛋白，如图 7-15 中的小峰(T^+)，但因其丰度很小难以准确测定。我们拟用糖苷内切酶酶切的方法切除糖链，获得蛋白部分的分子质量。*N*-连接糖链通常包含一个五糖核心，即三甘露糖

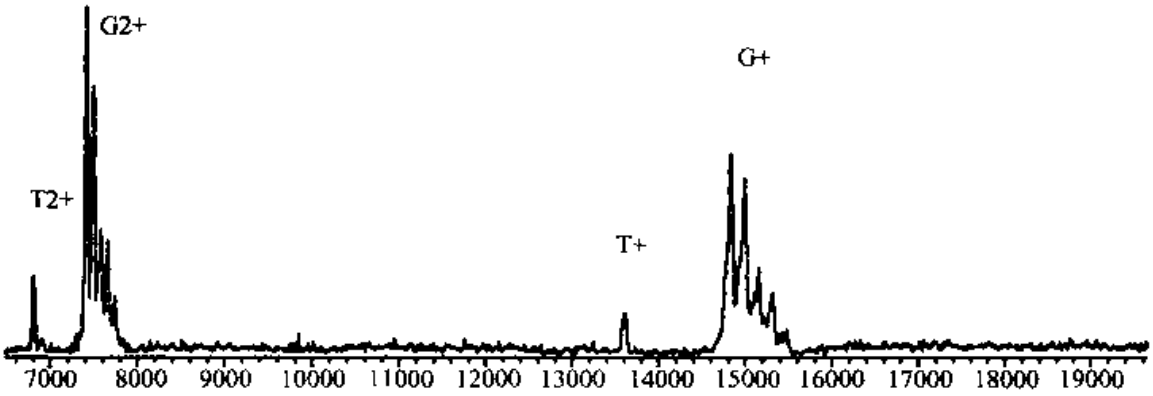


图 7-15　核糖核酸酶 B 的 MALDI-TOF-MS 图

五糖核心(图 7-11)。内糖苷键酶 F 切断 N -糖链五糖核心最内侧的 GlcNAc-GlcNAc 糖苷键,得到含一个 GlcNAc(M = 203Da)的肽链,从中减去 GlcNAc,可以计算出准确的肽链相对分子质量。肽链相对分子质量与糖蛋白平均相对分子质量之差为糖链的平均相对分子质量,进而求得糖蛋白的平均糖含量(图 7-16)。

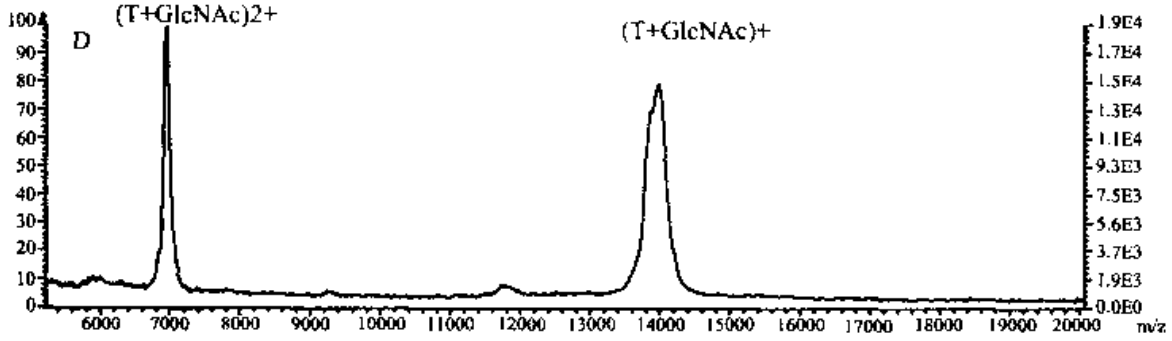


图 7-16　核糖核酸酶 B 用内糖苷键酶 F 酶切后的 MALDI-TOF-MS 图

(3)糖基化类型及糖基化位点的确定

糖链和肽链中的氨基酸残基通过共价键连接,用生物质谱法测定糖基化位点,一般是通过蛋白酶酶切与糖苷内切酶酶切相结合的方法,寻找质谱图上的峰位移,从位移峰的肽段序列中,寻找出可能连接糖链的氨基酸,从而确定其糖基化位点。这种方法摆脱了传统的先分解糖蛋白中糖链和蛋白、然后分别测定的方法,直接在 pmol 数量级上对糖蛋白进行分析,应用高灵敏度、高准确度的 ESI-MS 或 MALDI-TOF-MS,结合 HPLC 或 CE 分离技术以及网上数据库检索,可以快速、准确地确定自然界中或重组糖蛋白的糖基化位点。MALDI-TOF-MS 及 ESI-MS 这两种方法在准确度和灵敏度方面互补,在生物大分子分析方面起到了相辅相成的作用。

糖蛋白的糖链一般处于蛋白高级结构的外侧，因此直接切除糖链是可能的，对于二硫键较少的糖蛋白，直接进行蛋白酶切也是可行的。为了减少反应步骤，可以先酶切、再还原，这种方法简单，加入的反应试剂较少，不进行烷基化，可以避免反应过程中的脱盐步骤及烷基化不完全带来的不便。但欲使酶切反应完全，必须依靠二硫键的还原来实现，通常可以先将蛋白变性、还原，且烷基化，再用糖苷内切酶进行酶切，在质谱图上寻找发生位移的含糖肽段，这种方法由于反应完全，往往得到的肽段信息多，丰度高，但由于加入的反应试剂较多，会影响质谱分析，在 MALDI-TOF-MS 或 ESI-MS 前，往往需要脱盐。将核糖核酸酶 B 变性、烷基化并用 DTT 还原，然后以谷氨酸羧基酶(Glu-C)酶切(图 7-17)，许多肽段在内糖苷键酶 F 作用前后不发生位移(T1、T2、T3)，说明这些肽段不含糖，而在内糖苷键酶 F 作用后位移的肽段为含糖肽段(G4→T4)，我们将图 7-17 中的肽质量指纹图(PMF)数据进行数据库检索，证实 T4 肽段的序列为：T10 ～ 49：RQHMDSSTSAASSSNYCNQMMKSR NLT KDRCKPVNTFVHE，其中含有 N 糖链特异连接位点——天冬酰胺序列子(Asn-X-Thr/Ser)，由此我们确定了 34 位 Asn 为糖基化位点[44]。表 7-1 列出了数据库检索出的 Glu-C 蛋白酶酶切核糖核酸酶 B 肽段的 MALDI-TOF-MS 计算值、测定值及误差。

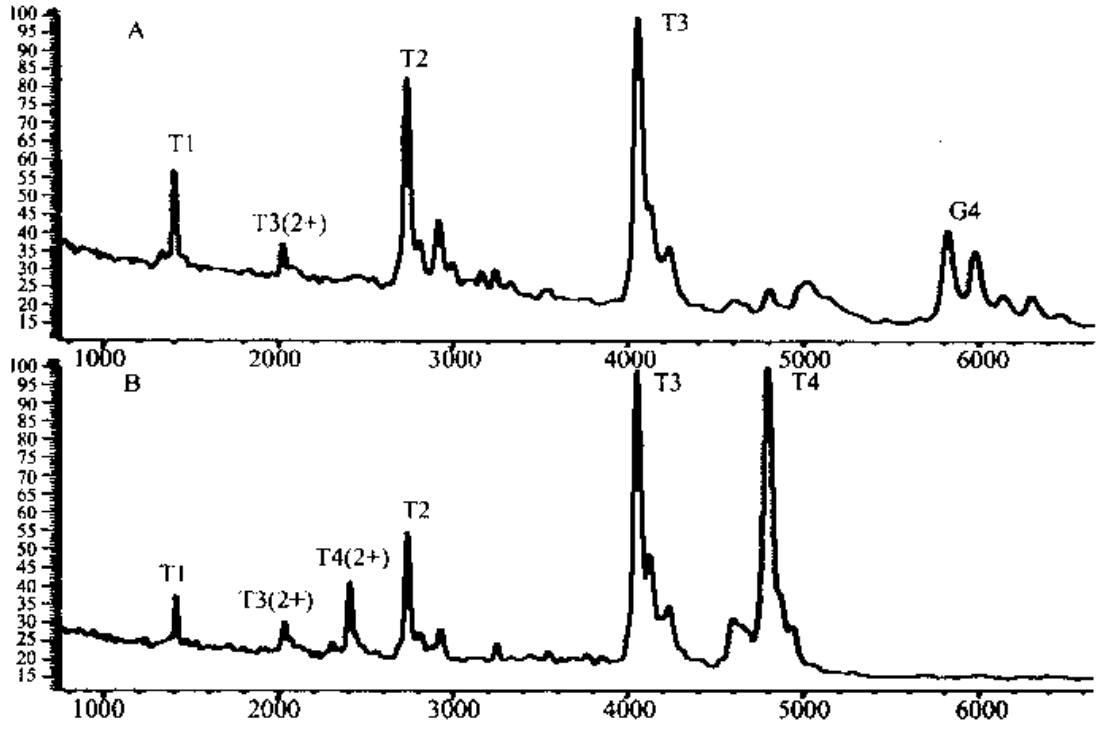


图 7-17　核糖核酸酶 B 的糖基化位点分析图

A. 还原烷基化并且 Glu-C 酶切。B. 在 A 的基础上加入内糖苷键酶 F。

表 7-1　Glu-C 蛋白酶酶切核糖核酸酶 B 肽段的 MALDI-TOF-MS 测定值

	肽　段	序　列	计算值	测定值	误差/%
T1	3～14	TAAAKFERQHMD	1405.5	1405.1	0.03
T2	87～111	TGSSKYPNCAYKTTQANKHIIVACE	2729.1	2727.6	0.09
T3	50～86	SLADVQAVCSQKNVACKNGQ	4048.5	4046.5	0.05
T4	10～49	RQHMDSSTSAASSSNYCNQMMKSR NLT KDRCKPVNTFVHE	4795.1	4792.5	0.05

唾液酸酶是用于将糖链外端修饰的唾液酸切除的一种特异性酶，而 *O* -聚糖酶则是切断 O-连接糖链 GalNAc-Thr/Ser 的一种酶，也是目前能够切断 *O* -连接糖链惟一的一种特异性酶，内糖苷键酶 F、肽- *N* -聚糖酶 F(PNGase-F)等都是作用于 *N* -糖链的糖苷酶，它们往往从糖链内侧切断糖链，专一性较好，是分析糖蛋白的有利工具。

2. LC-ESI-Q-TOF-MS

液相色谱电喷雾串联质谱通常以液相色谱进样系统将样品导入电喷雾离子源，在氮气的帮助下形成带多电荷离子的雾，经四极杆分离器获得一级质谱，进入气体碰撞活化室选择离子，以惰性气体与其碰撞产生碎片，进入飞行时间管道进行分离，得到二级质谱。这一离子化条件较其他方式更加缓和，因此无论是共价键或非共价键结合的化合物都可以检测，其谱图直接显示为多电荷峰，经转换得到单电荷分子离子峰。ESI 最优势的地方在于它是众多软电离方式中最"软"的一种，能与液相色谱(LC)或毛细管电泳(CE)串联，在 fmol 水平上测定混合物，如蛋白酶解产物，且无碎片离子峰。

(1)串联质谱测定糖蛋白的氨基酸序列

蛋白质氨基酸序列的测定虽然具有悠久的历史，但一直是沿用传统的 Edman 降解法测定 N 端氨基酸序列；近年来也以测定 C 端 1～2 个氨基酸来作为序列正确的依据。而对蛋白质内部序列，只能用蛋白酶切，HPLC 分离、收集肽片段，再分别对肽片段的两端测序的方法。这些方法不仅繁琐、样品需要量大，且存在局限性，比如 N 端测序中半胱氨酸丢失的问题、C 端测序中反应时间难以控制造成数据差错等，均给测定带来不便。电喷雾质谱的液相色谱导入系统使得它成为众多软电离方式中最"软"的一种，与灵敏度极高的串联质谱技术结合，可以选择肽质量指纹谱中的某一肽段进行序列分析，得到蛋白质某一中间片段的氨基酸序列，经网上数据库检索，确定其蛋白质的归属。对于已知蛋白，用串联质谱的方法，可以测定蛋白质氨基酸序列中间某一片段，从而判断其序列与理论序列是否一致，这一方法完全可以被延伸至糖蛋白的一级结构分析中来，可以将糖蛋白酶切后的肽段，包括含糖肽段，在串联质谱上直接选择离子进行碰撞活化，从而分析选择肽段的氨基酸序列，得到糖蛋白的一级结构信息。无论是重组糖蛋白还是胶上糖蛋白，都可以

用这种方法进行分析。我们对图 7-17 中的 T4 肽段应用英国 Micromass 公司的 Q-TOF2 型液相色谱电喷雾串联质谱仪进行测序(图 6-15),得到了核糖核酸酶 B 第 17～29 位氨基酸序列。

(2)糖链结构的分析

糖蛋白中的糖链形式各异、连接方式多样,传统的方法是将其进行氧化还原反应,再逐个分析。目前应用生物质谱的方法,大大改善了分析进程,通常以蛋白酶切、HPLC 分离得到含糖肽段,对含糖肽段可以直接用 ESI 串联质谱以及 MALDI-TOF-MS 的源后衰变技术(PSD),直接得到单糖碎片,从而确定糖链结构;也可用糖苷外切酶从糖链外端对糖链进行逐一切除并做质谱测定,不仅可以确定其连接方式,由所应用的糖苷外切酶的高度特异性,还可以判断糖链的立体结构。

随着电喷雾串联质谱仪器的发展,目前最直接的分析糖链结构的方法是应用串联质谱的母离子扫描(parent ion scan)技术,选择特定的糖链离子,经碰撞活化室用惰性气体将其打碎,从而推断糖链结构,比如在人血清转铁蛋白的糖链结构分析中[45],选择以 HPLC 分离的一个糖链,其分子离子双电荷峰$[M+2H]^{2+}$在电喷雾串联质谱中为 823.8,将其设定为母离子,以惰性气体与其碰撞产生碎片,从而分析其糖链结构(图 7-18)。

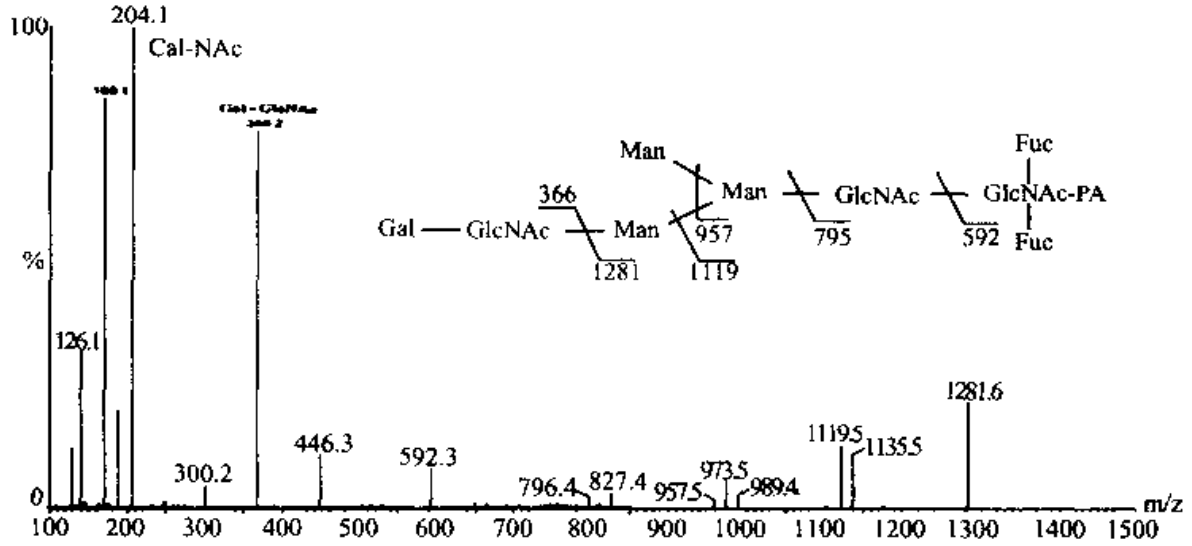


图 7-18 人血清转铁蛋白中糖链的电喷雾串联质谱图[45]

另外,也可应用中性丢失的方法,选择离子检测相应的肽段,例如图 7-19,将明胶酶 B 用胰酶及肽-N 聚糖酶降解,应用 MALDI-TOF-MS 测定核心岩藻糖(Δ)片段,由于糖链结构的复杂性,通常需要将质谱分析与其他辅助知识(如糖生物学方面的知识等等)相结合,共同确证糖链的结构。

虽然近两三年来应用质谱特别是串联质谱分析糖蛋白结构的报道越来越多,但绝大多数是应用 HPLC 技术对糖肽段进行分离、提取,这一步骤无疑繁杂,而且

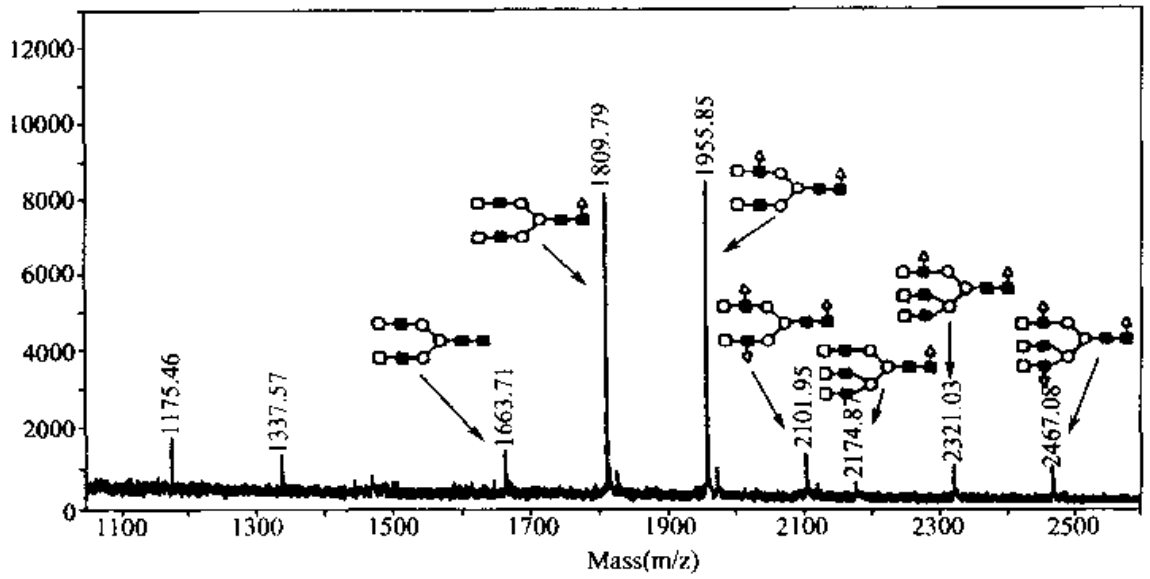


图 7-19　MALDI-TOF-MS 测定人明胶酶 B 的核心岩藻糖[46]

对样品需要量大，很难分析一些痕量样品(比如天然产品)。

3. 凝集素在糖蛋白分析中的应用

凝集素是一类糖结合蛋白，它能专一地识别某一特定结构的单糖或寡糖中特定的糖基序列而与之结合。因而凝集素不但可用于寡糖或糖复合物的分离纯化，也可用于糖链的结构分析，比如：伴刀豆球蛋白(ConA)对高甘露糖型 N-糖链有专一性；麦胚凝集素(WGA)却对杂合型 N-糖链有专一性；而亲和核心岩藻糖类糖链应选用兵豆凝集素等等。以核糖核酸酶 B 为例，我们用 ConA 提取了其高甘露糖

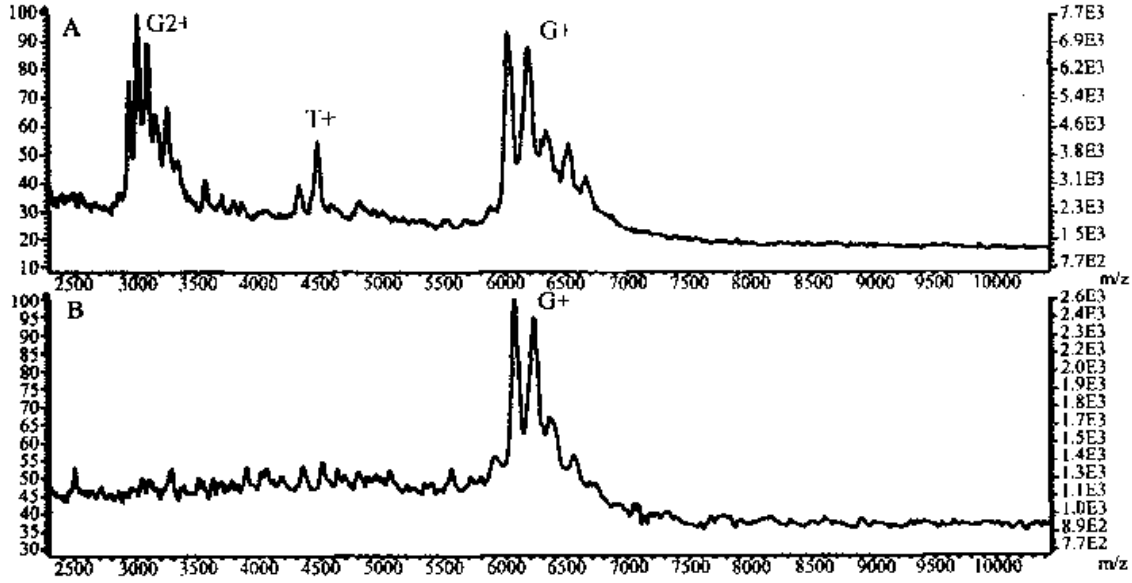


图 7-20　ConA 提取核糖核酸酶 B 的 Glu-C 酶切物的 MALDI-TOF-MS 图
A. 酶切肽图。B. ConA 提取糖肽段。

型糖链(图 7-20)。随着糖蛋白结构研究的日益深入,胶上酶切结合凝集素提取糖肽段,并在 MALDI 上直接检测的方法将被更加广泛地应用。

ESI 在糖蛋白或糖肽的离子化方面比 MALDI-TOF 更为成功,这可能是由于在 MALDI 中必须应用基质造成的,由于多糖的微不均一性、大分子的 N-连接糖链以及难以观察的多糖唾液酸化等等原因,使有些糖蛋白与基质难以发生衍生化。所以,在 MALDI-MS 图上,很难观察到大分子糖蛋白较好的峰形图。而 ESI-MS 的高分辨率以及使糖蛋白带多电荷检测,使糖蛋白在电喷雾质谱仪上的灵敏度及分辨率都大大改善,有时能够看到多糖结构的混合体,以及在质谱图上显示其微不均一性。

质谱可以测定不同糖基化形式中的单糖位点。通过碰撞诱导解离的特殊方式,这一技术还可以提供各个残基的序列、连接方式方面的信息。MALDI 及 ESI 这两种技术适合于高灵敏度地对大分子质量范围内的生物分子进行测定,因而它们可用于探索打碎糖复合物,目前已经获得了在对蛋白或多肽释放的非衍生化多糖结构分析方面的较好数据。MALDI-TOF 的源后衰变[47]或 CID-ESI-TOF 的源内碰撞-诱导解离[48]或 ESI-MS-MS[49]等技术可以打碎糖苷键。由于微孔及毛细管液相色谱的接口适合于被分离的糖基化形式的碰撞诱导碎片分析,目前电喷雾电离技术已成为蛋白糖基化结构分析的最佳选择,不同离子化技术与更好的检测方法结合是质谱的发展方向,在蛋白糖基化形式的鉴定方面,普遍认为这一方法比别的方法出色。

三、糖蛋白与蛋白质组学

1. 糖蛋白质组学的生物学意义

蛋白质组学的发展对整个生物体系的研究是一个新的突破,二维凝胶电泳(2-DE)的出现使细胞中蛋白质组分的变化可以一目了然地展现在人们面前。随着生物质谱以及网上数据库检索系统的发展,目前,人们已经能够快速、准确地以高通量方式在 fmol 数量级上从胶及膜上提取并鉴定蛋白质。众所周知,真核细胞的基因组与蛋白质组间的特征区别之一就是绝大多数基因产物是经过翻译后修饰的,其中糖基化是最重要的修饰方式。仅仅分析蛋白质的氨基酸序列,并不足以确定其特殊的功能,发生在蛋白表达之中或之后的修饰对蛋白质的功能影响非常大。因此,分析这些修饰对不同蛋白表达模型的理解是至关重要的,也就是说,蛋白质组的方法需要被延伸至分析细胞中糖蛋白。

糖基化过程是糖基转移酶活性的反映,糖基转移酶能够合成寡糖并将其附着于蛋白质,糖苷酶能够修整及降解多糖。许多丰度高的蛋白都是糖基化的,依糖基形式不同它们的等电点及相对分子质量不同,并且在蛋白质的糖基化过程中有高度的不均一性。糖蛋白的这些特殊性给其研究带来了困难。人们必须下大功夫去

研究这些糖基化变化如何对有机体产生生物学效应;同时,因为糖蛋白糖基组分固有的多样性能导致蛋白质在形状、电荷及体积等方面发生微妙的变化,这种变化有可能是暂时性的,或者在空间上修饰蛋白质的功能,这种多样性需要应用相关复杂的物理及化学技术对其组成、连接、分支、异构体及连接位点等进行确定,因此,研究多糖的分析方法通常被认为是化学工作者的职责。由于仅限于单糖残基结构及多糖连接方式的研究,通常碳水化合物分析化学被发展成为复杂多糖的结构测定,蛋白质中多糖的分析就是从这些技术中选择性地开发的。尽管如此,要想揭示有机体新陈代谢中糖基化的规律,仅仅对蛋白质中多糖结构定性是不够的,因为这些变化并不足以描述糖基赋予蛋白质的功能。此外,糖链的丢失对蛋白质活性的影响有时是观察不到的,比如一些重组糖蛋白在其天然糖基化被丢失后仍具有活性。普遍认为,糖基化参与蛋白的折叠,同时,糖基化的多样性在某种程度上决定了蛋白质功能的不规律性。

蛋白质组方法鉴定在疾病进程的初期疾病与正常组织之间所表现出来的蛋白质间的差别,这些蛋白质可作为诊断及控制疾病的工作焦点。由于血清通常用于临床诊断,同时易于获得,因此,有许多报道是关于在疾病过程中,血清中糖蛋白的糖基化变化的,我们推断在固体病变组织中表现出的糖基化变化也会有相类似的差异性。普遍认为二维凝胶电泳是展示蛋白表达过程中变化的首选方法,许多蛋白是同时存在的,而且观察到的差异与其表型差异是相关联的。蛋白质组并不需要以单个蛋白作为靶点,而是在任何可以改变生物体系的所有可能的网络路径中发掘其影响,所以蛋白质组要求对多种种属群体成员进行筛选,以揭示那些发生在蛋白表达中与个体水平相类似的群体水平上的变化。如果我们认为一个糖蛋白对疾病的表型变化有影响,那么它就可以被用作疾病的诊断标记,或者直接用凝胶电泳对其进行分析;再者,也可利用鉴定蛋白糖基化变化的特殊标志方面的知识,衍生出其他分析方法对其进行分析。目前,2-DE 已经成功地用于鉴定糖蛋白差异,已报道众多的血清糖蛋白作为肾细胞癌的标记物[50];同时,糖基化差异也被用于构建多糖类癌症疫苗,以特殊的多糖与合成多肽结合,用于免疫治疗的不同策略已在临床试用[51,52],目前,这种方法已经成为这一领域的研究兴奋点。

许多发生在目标蛋白糖基化中的变化,有的是由于糖基化程度不同而导致分子量的变化,有的是由于相应的唾液酸化而引起等电点变化,同时,多糖的磷酸化及硫酸化也会使糖蛋白的等电点发生变化。糖蛋白的这些固有属性使它们在 2-DE 上理想地排列,这一阵列非常适合检测糖基化过程中的特征性变化,目前还没有其他技术能够以这种形式展示糖蛋白的不同。

2. 经分离的胶上糖蛋白分析

由于纯化蛋白所提供的量是有限的,所以用一维或二维凝胶电泳分离蛋白从而实现糖基化结构变化检测的方法受到了挑战。质谱(包括基质辅助激光解吸附

飞行时间质谱和电喷雾串联质谱)及毛细管区带电泳(CIE),是两种用于解决出现在糖蛋白异构形中糖链结构的不均一性问题的最有前途的技术。糖蛋白的物理及化学特性,给应用现有蛋白质组技术对其进行分析造成了困难,糖链的连接(特别是大型的N-连接糖链),由于其庞大的相对分子质量及亲水性,使蛋白质的特性千差万别,为了分析 2-DE 分离的糖蛋白,蛋白质组进程需要在许多方面加以优化。

首先,在实验中发现,将纯化的糖蛋白在一维固相 pH 梯度(IPG)胶条上上样时,由于被动再水化现象,使进入胶条的样品量很低,利用"杯状上样法"可以使糖蛋白进入 IPG 胶条的量大大增加,其原因只能如上推断,但是它确实能够反映多糖对蛋白质表面积大小、电荷等方面的影响,同时,它也有助于揭示糖蛋白物理特性的不同。在"考染"过程中,人们发现,糖蛋白的染色非常不理想,或者根本染不上色。为了对 2-DE 胶上蛋白质混合物阵列中的糖蛋白着色,最好的办法是染色其特殊的多糖部分,即应用灵敏的凝集素及半抗原识别体,目前市场上已经出现了许多基于糖的高碘酸盐氧化作用,以及夹层抗体或生物素/链霉亲合素等显色方法发展起来的产品(Bio-Rad 多糖检测试剂盒,Boehringer-Mannheim DIG 检测试剂盒)。也可以用荧光替代颜色,对氧化态糖的显色方法加以改善,从而使 PVDF 膜上的糖蛋白快速、灵敏地显现。在用 2-DE 分析人乳腺癌标本过程中[53],使用 HPA(Helixpomatia)凝集素键合技术,以化学发光法将糖蛋白显色,从而得到标本中糖蛋白的 2-DE 谱图(图 7-21)。

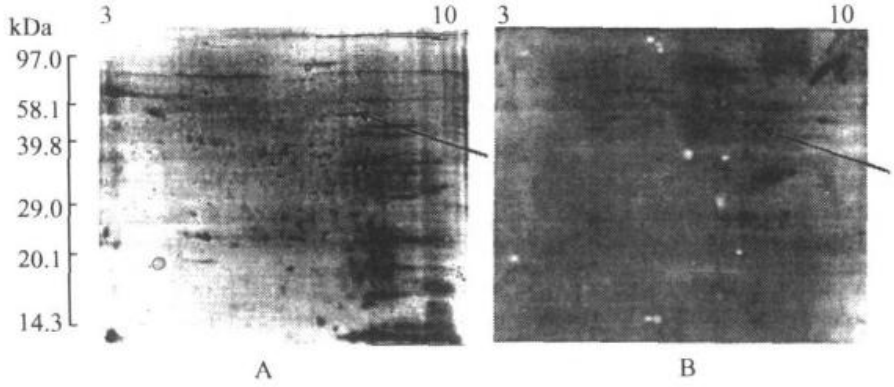


图 7-21　人乳腺癌标本的 2-DE 图[52]

A. 银染蛋白图。B. 同样的样品以 PVDF 膜转印,用 HPA 染色并用 ECL 显色。

其次,造成在 2-DE 胶上单个糖类型个别分离不明显的原因是由于多糖的不均一性,这种在等电点上的差异有时是因为唾液酸化的不同。但是,在用质谱分析人 α1-抗胰蛋白酶的 2-DE 胶上蛋白的 N-连接糖链时发现,在每个糖基形式中,中性糖的唾液酸化有相同的比率,造成等电点降低的直接原因是由于在不同的点上,糖基化是不均一的[54]。

ESI 与 MALDI 的快速发展对蛋白质分析的促进是革命性的,而对多糖的分析也存在着同样的潜力,N-连接糖链质量(包括多肽部分)通常大于 3500Da,在

MALDI 上难以看到，加之以不均一化的趋势，经常只能看到一个宽的信号峰。O-连接多糖在总的趋势上比 N-连接多糖分子要小，胶上酶解后，在质谱上有时可以看到糖蛋白的胰酶糖肽碎片。将胶上电泳分离并用 PVDF 膜斑点化的糖蛋白进行肼解或碱性β消除释放糖链，也可对 O-连接多糖进行定性分析。通过毛细管区带电泳技术可以观察糖蛋白等电点及结构上的微小差异，应用质谱技术可以对其进行定性分析，但是被分析物需要有一定的浓度。首先，我们对蛋白质高通量分析方法的选择，是应用灵敏、快速、简单的 MALDI 自动方法从一个已知基因组对蛋白进行鉴定，这种方法对于一些小的考染蛋白斑点的胰酶降解碎片也适用。一旦蛋白在数据库上鉴定，许多被修饰的蛋白的微小特性也就相应显现，接着用 ESI-MS-MS 技术作为精细结构分析的工具，鉴定被修饰的特殊蛋白，为了彻底确定糖基化性质，需要同时分析其微不均一性及宏观差异。

糖蛋白质组质谱常规应用方面的问题在于，对于 MALDI-TOF-MS，适合于大分子亲水性糖肽的特殊基质有待于进一步开发，同时，应用 MALDI 分析时，糖肽的上游富集的选择是很有帮助的；在 ESI-MS-MS 方面，糖肽虽然可以在 LC 上分离，但糖肽上的多糖碎片难以被解释，检测灵敏度需要进一步提高，或者，需要寻找一个易离子化基团对目标糖进行衍生化，其化学反应必须完全且简单、同时能够适应高通量的分析，使自动化进程得以加速。

尽管如此，对蛋白质上糖类型的完全结构分析，其要求远远超过仅仅对糖进行相对分子质量测定。与用于肽质量指纹谱中的氨基酸质量相比，许多单糖残基有着相同的相对分子质量，同时，连接方式、序列以及空间构型难以用质谱测定。但是，我们也应该看到，由于各种单糖在其结构和框架范围内有一定的守恒性，使糖蛋白的分析简化。因此，可以通过对一些已知的特殊组织或种属方面的知识，来约束可能的具有某种特定质量的多糖结构。

3. 基因工程产品中的糖蛋白分析

糖分析最初是以复杂的、大相对分子质量的多糖的化学及物理结构测定为焦点的。这些技术从前非常适于对于那些相对分子质量较小、结构较简单的多糖的分析。质谱仪的发展，加上许多特殊的外切糖苷酶的纯系化，使我们能够分析许多糖蛋白上寡糖链的位点、连接方式及序列。由于经纯化过的被表达的重组糖蛋白在样品量的提供方面占有优势，所以通过对它们的分析，使得绝大多数分析方法得以优化。相对而言，分析 2-DE 分离的糖蛋白所需的样品量日益减小(通常为 10 pmol 或 100ng)。

由重组 DNA 生产蛋白质的方法是生物技术产业的一个延伸领域。许多治疗及诊断蛋白质都是糖基化的，因此糖基化形式的快速分离及鉴定就十分必要。由于多糖链的变化能够诱导许多不良的免疫学反应，导致越来越多的需求对糖蛋白的多糖部分进行严格的鉴定。对于细胞系和培养条件的依赖，使表达体系不可能

完全控制蛋白的糖基化，众多的结构差异由此产生。通常可以用控制发酵过程中的条件的方法，初步监测不同糖基结构是怎样产生或者何时产生的；同时，用液相色谱法监测批次间差异及下游进程；电泳技术由于其高通量、对比性分析及分辨率等方面的优势[55]，如等电聚焦、毛细管电泳及二维凝胶电泳等，都被用做快速常规分析及产品监测。

糖基化的出现与缺失也能调节一些重组蛋白的功能，比如：人粒细胞集落刺激因子[56]、白细胞介素 2[57]、甲状腺过氧化物酶[58]及 α1 -酸性糖蛋白[59]都是要求具有体内活性的糖基化重组蛋白。其他一些重组蛋白天然就存在糖基化，如血小板生成素[60]及人生长激素[61]，它们甚至在无糖基化时也有活性。我们对重组人促红细胞生成素(rhEPO)的糖含量进行了 MALDI-TOF-MS 分析(图 7-22)，根据酶切前后的质量差可以计算出唾液酸含量及总糖含量，其中唾液酸含量为 8.6%，总糖含量为 34.4%。

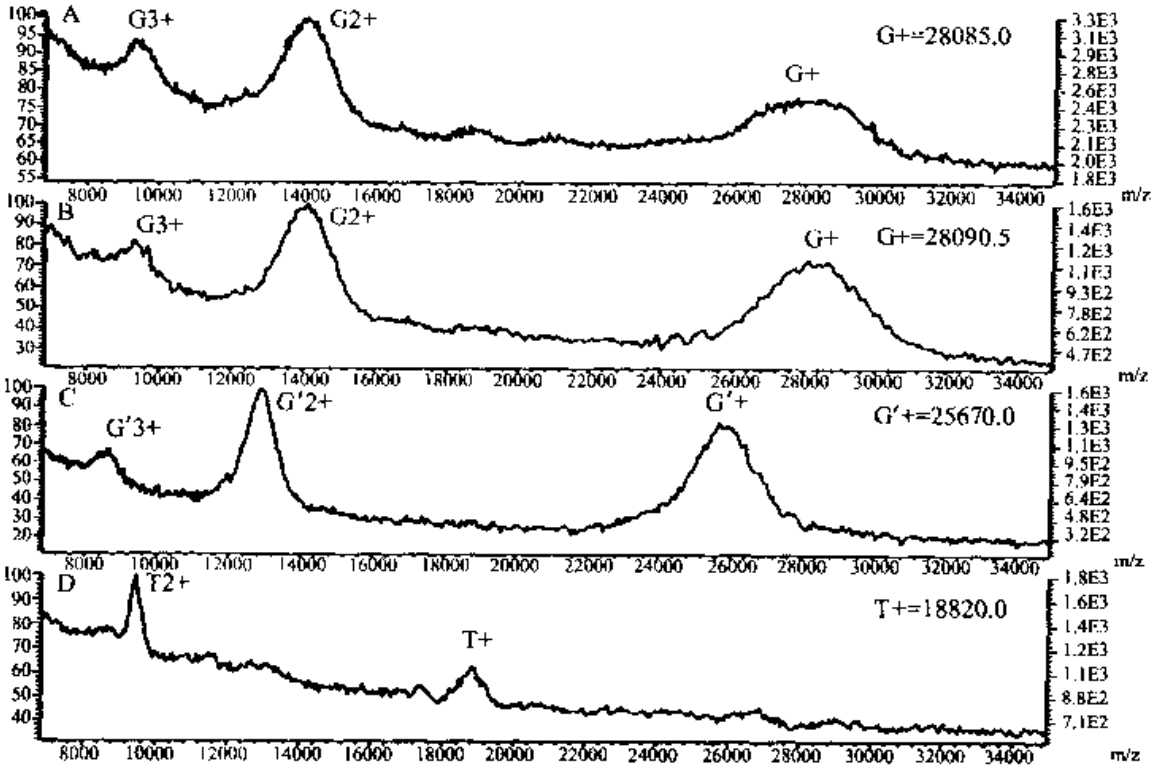


图 7-22　EPO 分子质量的 MALDI-TOF-MS 测定

A. EPO 分子质量。B. O-聚糖酶酶切。C. 唾液酸酶酶切。D. 内糖苷键酶 F 酶切。G+、G2+、G3+ 为糖蛋白 EPO 分子离子峰及双电荷、三电荷峰，T+、T2+ 为不含糖的 EPO 分子离子峰及双电荷峰。

最后，基因组及蛋白质组领域所获得的大量数据使生物信息学领域受益匪浅，SWISS-PROT 及类似的许多数据库给生物化学家们提供了信息的贮藏地，进而也需要类似于它的完整的数据库及信息工具适应糖基化数据的翻译，以适应糖分析

的发展。将糖肽或被释放的糖的质量分析转化成单糖构架的分析，结合生物学以及化学知识，同时通过设计特殊的糖苷外切酶阵列，可以检测可能的结构。根据糖质量指纹谱还可以推测序列，综合糖蛋白中大量有价值的知识，用以分析从糖蛋白质组分析中获取的数据。

结语

蛋白质的翻译后修饰，包括糖基化，对于发生在生物学体系中的新陈代谢具有控制作用，多糖结构中存在着可利用的广泛的差异性，这些差异性对于新陈代谢的调节细致入微。2-DE 展示的糖蛋白阵列为我们提供了一个谱图，由它我们可以探索疾病发生、发展过程中的变化。重组糖蛋白由于其能够提供足够的样品量，应用质谱以及相关技术的分析工具已经能够较完善地鉴定其结构。目前，要求这些可达到 pmol 灵敏度的检测方法能够像蛋白质组鉴定的一些方法一样，适应高通量的糖蛋白的鉴定与定性。随着各种软电离方式的出现，质谱仪器已经被广泛应用于生物样品的分析，加之各种分离手段如 HPLC、CE 与串联质谱的结合，以及 MALDI-TOF-MS 中 DE、PSD 技术的应用，使质谱仪器更加多样化，满足生命科学研究中对于生物大分子结构信息的需求。信息高速公路的开通，Internet 网上数据库可以瞬间提供大量的信息，以肽质量指纹谱结合网上数据库检索确定蛋白质一级结构。凝集素提取糖肽片段直接 MALDI-TOF-MS 分析、二维凝胶色谱(2D-Gel)分离大量混合蛋白并胶上酶切技术、靶上直接酶切质谱分析等快速、准确、灵敏的新方法，已经使蛋白质的结构分析进入了自动化时代，相信这些技术同样适用于糖蛋白的结构分析。从已开发的数据形式我们可以获得延伸的信息，这将依赖于改进的、更加灵敏的方法学方面的进展，以及生物信息特别是糖生物信息工具的构建。在不久的将来，人们将能够得到糖蛋白更多的结构信息，从而揭示它参与生命活动的奥秘。

参考文献

1 Hitto K, James E B, Martin F. Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis. Proteomics, 2001, 1: 194～199

2 敖世洲. 蛋白质可逆磷酸化对细胞活动的调节. 见：国家高技术研究发展计划生物技术领域战略研讨会论文集. 上海：上海科学技术出版社, 1994

3 Scott D P, Ruedi A, David R G. Mass spectrometry-based methods for protein identification and phosphorylation site analysis. In: Pennington S R and Dun, M J eds. Proteomics from protein sequence to function. New York: Springer-Verlag New York Inc, 2001, 87～130

4 [美]奥斯伯, 布伦特 等. 精编分子生物学实验指南(颜子颖, 王海林译, 金冬雁校). 北京：科学出版社, 1998, 683～695

5 Moese S, Selbach M, Zimny-Arndt U, Jungblut P R, Meyer T F, Backert S. Identification of a tyrosine-phosphorylated 35 kDa carboxy-terminal fragment (p35CagA) of the Helicobacter pylori CagA protein in

phagocytic cells: processing or breakage? Proteomics, 2001, 1(4): 618～629

6 Backert S, Muller E C, Jungblut P R, Meyer T F. Tyrosine phosphorylation patterns and size modification of the Helicobacter pylori CagA protein after translocation into gastric epithelial cells. Proteomics, 2001, 1(4): 608～617

7 Is'harc H, Watling D, Kerr I M. Phosphotyrosine profiling to identify novel components of interferon and interleukin 6-family cytokine signalling. Proteomics, 2001, 1(6): 767～772

8 Soskic V, Gorlach M, Poznanovic S, Boehmer F D, Godovac-Zimmermann J. Functional proteomics analysis of signal transduction pathways of the platelet-derived growth factor beta receptor. Biochemistry, 1999, 38(6): 1757～1764

9 Pandey A, Podtelejnikov A V, Blagoev B, Bustelo X R, Mann M, Lodish H F. Analysis of receptor signaling pathways by mass spectrometry: identification of vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. Proc Natl Acad Sci U S A, 2000, 97(1): 179～184

10 Steen H, Kuster B, Fernandez M, Pandey A, Mann M. Tyrosine phosphorylation mapping of the EGF receptor signaling pathway. J Biol Chem, 2001

11 Csar X F, Wilson N J, McMahon K A, Marks D C, Beecroft T L, Ward A C, Whitty G A, Kanangasundarum V, Hamilton J A. Proteomic analysis of macrophage differentiation. p46/52(Shc) Tyrosine phosphorylation is required for CSF-1-mediated macrophage differentiation. J Biol Chem, 2001, 276(28): 26211～26217

12 McLachlin D T, Chait B T. Analysis of phosphorylated proteins and peptides by mass spectrometry. Curr Opin Chem Biol, 2001, 5(5): 591～602

13 Muszynska G, Andersson L, Porath J. Selective adsorption of Phospho-proteins on gel-immobllized ferric chelate. Biochemistry, 1986, 25: 6850～6853

14 Betts J C, Blackstock W P, Ward M A, Anderton B H. Identification of phosphorylation sites on neurofilament proteins by nanoelectrospray mass spectrometry. J Biol Chem, 1997, 272(20): 12922～12927

15 Sickmann A, Meyer H E. *Phosphoamino acid analysis*. Proteomics, 2001, 1(2): 200～206

16 Stensballe A, Andersen S, Jensen O N. Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. Proteomics, 2001, 1(2): 207～222

17 Xhou W, Merrick B A, Khaledi M G, Tomer K B. Detection and sequencing of phosphopeptides affinity bound to immobilized metal ion beads by matrix-assisted laser desorption/ionization mass spectrometry. J Am Soc Mass Spectrom, 2000, 11(4): 273～282

18 Oda Y, Nagasu T, Chait B T. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. Nat Biotechnol, 2001, 19(4): 379～382

19 Zhou H, Watts J D, Aebersold R. A systematic approach to the analysis of protein phosphorylation. Nat Biotechnol, 2001, 19(4): 375～378

20 Larsen M R, Sorensen G L, Fey S J, Larsen P M, Roepstorff P. Phospho-proteomics: evaluation of the use of enzymatic de-phosphorylation and differential mass spectrometric peptide mass mapping for site specific phosphorylation assignment in proteins separated by gel electrophoresis. Proteomics, 2001, 1(2): 223～238

21 Zhang X, Herring C J, Romano P R, Szczepanowska J, Brzeska H, Hinnebusch A G, Qin J. Identification of phosphorylation sites in proteins separated by polyacrylamide gel electrophoresis. Anal Chem, 1998, 70(10): 2050～2059

22 Liao P C, Leykam J, Andrews P C, Gage D A, Allison J. An approach to locate phosphorylation sites in a *phosphoprotein*: mass mapping by combining specific enzymatic degradation with matrix-assisted laser desorption/ionization mass spectrometry. Anal Biochem, 1994, 219(1): 9～20

23 Zugaro L M, Reid G E, Ji H, Eddes J S, Murphy A C, Burgess A W, Simpson R J. Characterization of rat brain stathmin isoforms by two-dimensional gel electrophoresis-matrix assisted laser desorption/ionization and electrospray ionization-ion trap mass spectrometry. Electrophoresis, 1998, 19(5): 867～876

24 Metzger S, Hoffmann R. Studies on the dephosphorylation of phosphotyrosine-containing peptides during post-source decay in matrix-assisted laser desorption/ionization. J Mass Spectrom, 2000, 35: 1165－1177
25 Muller D R, Schindler P, Coulot M, Voshol H, van Oostrum J. Mass spectrometric characterization of stathmin isoforms separated by 2D PAGE. J Mass Spectrom, 1999, 34: 336－345
26 Asara J M, Allison J. Enhanced detection of phosphopeptides in matrix-assisted laser desorption/ionization mass spectrometry using ammonium salts. J Am Soc Mass Spectrom, 1999, 10(1): 35－44
27 Ma Y, Lu Y, Zeng H, Ron D, Mo W, Neubert T A. Characterization of phosphopeptides from protein digests using matrix assisted laser desorption/ionization time of flight mass spectrometry and nanoelectrospray quadrupole time-of-flight mass spectrometry. Rapid Commun Mass Spectrom, 2001, 15(18): 1693－1700
28 Hunter A P, Games D E. Chromatographic and mass spectrometric methods for the identification of phosphorylation sites in phosphoproteins. Rapid Commun Mass Spectrom, 1994, 8(7): 559－570
29 Schlosser A, Pipkorn R, Bossemeyer D, Lehmann W D. Analysis of protein phosphorylation by a combination

45 Eric A, Noriko T, Yoshinori T, et al. *N*-glycan patterns of human transferrin produced in Trichoplusia ni insect cell: Effects of mammalian galactransferase. Glycobiology, 2000, 10(8): 837～847
46 Proceedings of the 46th ASMS conference on mass spectrometry and allied topics. Orlando Florida. 1998, 3
47 Naven T J, Harvey D J, Brown J Critchley G. Fragmentation of complex carbohydrates following ionisation by matrix-assisted laser desorption with an instrument fitted with time-tag focusing. Rapid Commun Mass Spectrom, 1997, 11: 1681～1686
48 Packer N H, Harrison M J. Glycobiology and proteomics: is mass spectrometry the holy grail? Electrophoresis, 1998, 19: 1872～1882
49 Oxley D, Munro, S L, Craik D J, Bacic A. Structure and distribution of N-glycans on the S7-allele stylar self-incompatibility ribonuclease of Nicotiana. J Biochem (Tokyo), 1998, 123: 978～983
50 Delahunt B, Maxwell J, Moroni-Rawson P, Renner R M, Nacey J N, Jordan T W. Serum protein variation in the patients with renal cell carcinoma. J Urol Pathol, 1996, 5: 97～107
51 Kudryashov V, Kim, H M, Ragupathi G, Danishefsky S J, Livingston P O, Lioyd K O. Immunogenicity of synthetic conjugates of Lewis(y) oligosaccharide with proteins in mice: towards the design of anticancer vaccines. Cancer Immunol Immunother, 1998, 45: 281～286
52 Lo-Man R, Bay S, Vichier-Guerre S, Deriaud E, Cantacuzene D, Leclere C. A fully synthetic immunogen carrying a carcinoma-associated carbohydrate for active specific immunotherapy. Cancer Res, 1999, 59: 1520～1524
53 Miriam V, Dwek Heidi A, Ross Anthony J. C. Leathem. Proteome and glycosylation mapping identifies post-translational modifications associated with aggressive breast cancer. Proteomics, 2001, 1: 756～762
54 Packer N H, Lawson M A, Jardine D R, Sanchez J C, Gooley A A. Analysing glycoproteins separated by two-dimensional gel electrophoresis. Electrophoresis, 1998, 19: 981～988
55 Taverna M, Tran N T, Merry T, Hovarth E, Freeier D. Electrophoretic methods for process monitoring and the quality assessment of recombinant glycoproteins. Electrophoresis, 1998, 19: 2572～2594
56 Hoglund M. Glycosylated and non-glycosylated recombinant human granulocyte colony-stimulating factor (rhG-CSF) What is the difference? Med Oncol, 1998, 15: 229～233
57 Voshol H, Dullens H F, Otter W D, Vliegenthart J F. The glycosylation profile of interleukin-2 activated human lymphocytes correlates to their anti-tumour activity. Anticancer Res, 1996, 16: 155～159
58 Fayadat L, Niccoli-Sire P, Lanet J, France J L. Human thyroperoxidase is largely retained and rapid degraded in the endoplasmic reticulum. Its N-glycans are required for folding and intracellular trafficking. Endocrine, 1998, 139: 4277～4285
59 Shiyan S, Bovin N V. Carbohydrate composition and immunomodulatory activity of different glycoforms of alphal-acid glycoprotein. Glycoconj J, 1997, 14: 631～638
60 Jagerschmidt A, Fleury V, Anger = leroy M, Thomas C, Agnel M, O'Brein D P. Human thrombopoietin structure-function relationships: identification of functionally important residues. Biochem J, 1998, 333: 729～734
61 Becker G W, Hsiung H M. Expression secretion and folding of human growth hormone in Escherichia coli. Purification and characterization. FEBS, 1986, 204: 145～150

（蔡　耘　王京兰）

第八章　定量蛋白质组学研究技术

如果把 O'Farrell[1]在 1975 年运用二维凝胶电泳技术对 *E. coli* 中总蛋白进行分离作为研究蛋白质组学技术的开始，蛋白质组学研究技术也不过才仅仅几十年的历史。但是，它却得到了很大的发展，包括 20 世纪 80 年代固相化 pH 梯度的发明和完善，生物质谱技术在鉴定蛋白质中的成功应用以及计算机技术在分析二维凝胶电泳图谱的应用。特别是在"蛋白质组"[2]的提出到现在这几年，新技术更是层出不穷。如多维色谱-质谱联用技术[3]、蛋白质荧光染色技术[4,5]、蛋白质芯片技术、同位素标记技术[6]、同位素亲和标签技术[7]等。这些新技术的使用突破了常规 2DE-MS 分析途径本身所固有的局限性，提高了分析的重复性、灵敏度、准确性。因此，有人提出了"定量蛋白质组学(quantitative proteomics)"概念[6,8,9]。定量蛋白质组学，顾名思义，就是把一个基因组表达的全部蛋白质或一个复杂的混合体系中所有的蛋白质进行精确的定量和鉴定的一门学科[8]。

传统的对蛋白质组研究的途径是分离-鉴定。对一个蛋白质混合体系，首先用二维凝胶电泳进行分离，然后对蛋白质点进行染色，最后通过比较单个蛋白点染色强度，对其进行定量分析。在这个过程中，理想的染色方法必须有以下几个特点：①具有良好的线性范围，也就是说，染色的强度与蛋白质的量成正比；②具有好的通用性，即对不同蛋白质具有相同的染色能力；③具有高的灵敏度，只有这样它才能检测出细胞中低拷贝蛋白质；④染色不影响后续的鉴定过程，即它不影响蛋白质分子在质谱仪中的离子化过程[4]。但是，目前大多采用考马斯亮蓝染色法或银染法，都不能完全符合上述的几点特征，也就不能应用于定量蛋白质组学研究。近年来，随着图像分析系统的发展，荧光染色技术更多地用于蛋白质组学研究。与考马斯亮蓝染色法或银染法相比，它能更好满足于上述几点要求，从而能够应用于定量蛋白质组学研究(后面将有详细的介绍)。在传统的蛋白质组研究中，质谱只应用于对未知蛋白点的鉴定，而不能应用于对蛋白质点的定量分析，这是因为不同的蛋白质或多肽具有不同的离子化能力，所以不能够通过质谱图对蛋白质或多肽进行定量分析。但是，Oda[10]等人提出的稳定同位素代谢标记的方法和 Gygi[7]等人提出的同位素亲和标签技术巧妙解决了这个问题。此外，还有蛋白质芯片技术，也可用于定量蛋白质组学研究。接下来将详细介绍这些技术。

第一节　利用荧光染料进行定量的蛋白质组分析技术

截止到现在，虽然有多种新技术应用于蛋白质混合物的分离，如多维色谱技

术、毛细管电泳技术等。但是没有一项新技术能比得上二维凝胶电泳的分辨能力，这使得二维凝胶电泳技术仍然是蛋白质组研究中首选的分离技术。在这个过程中，合适的显色技术是定量蛋白质的关键。Rabilloud[11]概述了应用于二维凝胶电泳分离的蛋白质显色方法，包括放射性标记方法、有机染料方法、荧光染料方法、金属离子还原法、基于对盐离子结合能力不同的负染法。此外，他还根据二维凝胶的三个过程，即等电聚焦前、等电聚焦与 SDS－聚丙烯酰胺电泳之间、SDS－聚丙烯酰胺电泳后，讨论了适用于各个过程的染色方法。但是，目前大多数实验室都采用 SDS－聚丙烯酰胺电泳后对分离的蛋白质进行染色的方法。在这个过程中，除放射性标记以外，其他方法均可用。前几年，由于没有合适的检测经过荧光染料染色的二维凝胶的设备，人们大多采用考马斯亮蓝染色法或银染法。对于考马斯亮蓝法来说，它虽然有好的线性范围、通用性且不干涉后续的质谱鉴定，但是它的灵敏性太低，所以不适合于蛋白质组研究。对于银染法来说，虽然有好的灵敏性，但它的线性范围太窄，且影响后续的质谱鉴定。由于上述所讲的几点原因，考马斯亮蓝染色法或银染法不能符合人们的要求，人们把目光投向了荧光染料，而此时，有厂家推出了适合于分析二维凝胶的荧光扫描仪，使得人们能够对经过荧光染色后的二维凝胶进行分析。

适用于检测蛋白质的荧光染料有两种：一种是能与蛋白质共价结合的荧光染料，如花青染料。由于这种结合是多位点的，因此它降低了蛋白质的溶解度。所以在运用共价结合染料来标记蛋白质时，染料的用量很少，这使得其灵敏度降低；此外，在二维凝胶电泳过程中，由于有大量未标记蛋白质存在，给后续的蛋白质点的切割和分析带来了困难。另一种是非共价结合荧光染料，也有两种不同类型：一种是荧光染料探针与蛋白质没有特定的亲和力，在疏水性环境中和亲水性环境中性质不同，只在疏水性环境中才会产生荧光，如苯乙烯基染料（如 Sypro Orange、Sypro Red 和 Sypro Ruby 等）和尼罗河红（Nile Red）等；另一种是荧光染料与蛋白质有特定的亲和力，如一种钌螯合剂（RuBPS）[4]。Lopez[12]等比较了在二维凝胶电泳以后分别用酸性硝酸银染色法与荧光染料 Sypro Ruby 在灵敏度、动力学线性范围和胶内酶解后运用基质辅助激光解析电离质谱回收肽段三方面能力的差异。通过对标准蛋白混合物进行 SDS－PAGE 一维电泳分离和对鼠纤维原细胞裂解物进行二维电泳分离后，进行胶内酶切、质谱鉴定来分析。在鼠纤维原细胞裂解物二维凝胶电泳图上，用银染法得到了 1290 个点，而用 Sypro Ruby 荧光染色法得到了 1063 个点，可见银染法灵敏度要比 Sypro Ruby 荧光染色法高一点。但对标准蛋白质胶内酶切和质谱分析后，对于含量大的磷酸化酶 b 蛋白（大于 500fmol），质谱图上肽段的回收率是一样的，而对于含量较少的磷酸化酶 b 蛋白（75fmol），经过银染后的蛋白质的肽质量指纹图上没有匹配的肽段，但经过 Sypro Ruby 荧光染色的蛋白质的肽质量指纹图上有 2 个匹配的肽段。对鼠纤维原细胞裂解物二维凝胶电泳图上一个高丰度点进行胶内酶切和质谱分析后，发现经过 Sypro Ruby 荧光染色的

蛋白质的肽质量指纹图要明显好于银染后的蛋白质的肽质量指纹图。他们还比较了两种染色方法的动力学线性范围，发现 Sypro Ruby 荧光染料的线性范围是酸性银染法的 700 倍。从上述比较可以看出，运用荧光染料染色的方法比银染法更适合于定量蛋白质组学研究。此外，Rabilloud[4]等人比较了两种不同类型的非共价结合荧光染料——Sypro Ruby 和 RuBPS 对胶内蛋白的染色和分析，发现这两种染料在灵敏度、线性范围、肽质量指纹图干扰方面没有大的差别。最近 Amersham 公司提出了荧光双向差示凝胶电泳(fluorescence two-dimensional differential gel electrophoresis, 2-D DIGE)技术，也就是运用不同的荧光染料分别标记不同的蛋白质样品，然后将它们等量混合起来，在单一的双向胶上进行分离和检测，按照它们所发出荧光的不同，来比较在不同细胞状态下蛋白质量的变化。由于它克服了在二维凝胶电泳过程中不同凝胶间重复性差的问题，定量更加准确。但是由于它是在等电聚焦前对蛋白质进行标记，所以必须应用与蛋白质共价结合的荧光染料，因此带来了一系列的问题，如上所述。其操作过程如图 8-1 所示。Tonge R[13]等运用 2-D DIGE 技术对正常鼠肝组织与扑热息痛处理的鼠肝组织的蛋白质组进行定量分析。详细情况参阅参考文献 13。

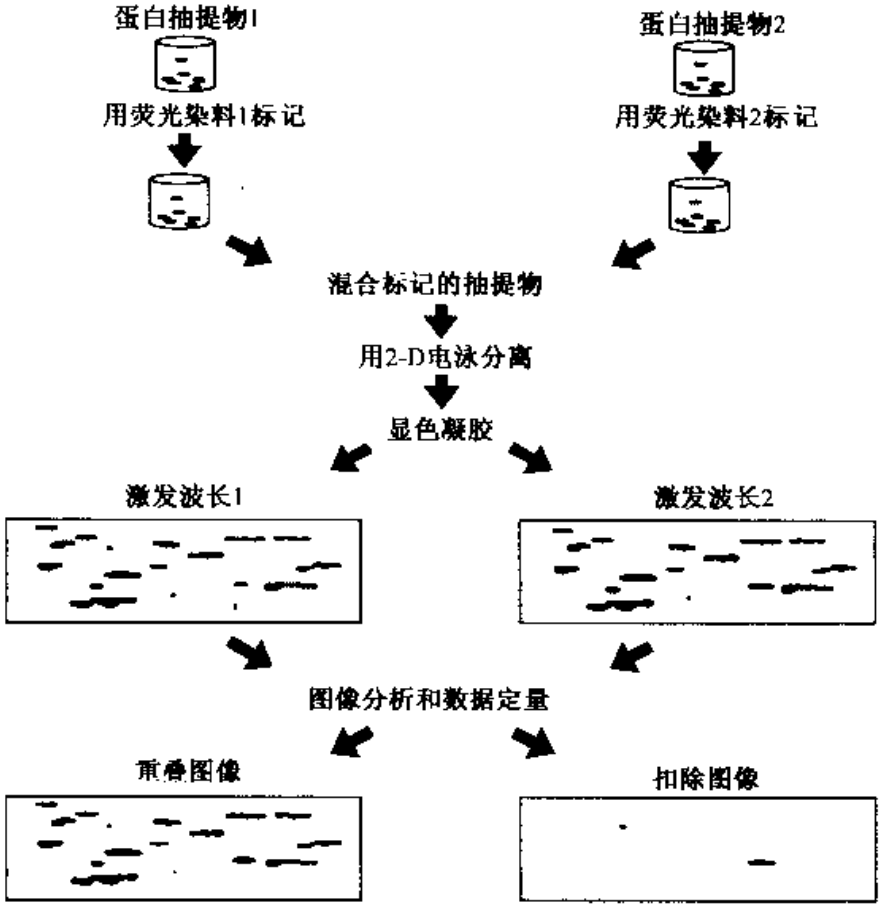


图 8-1　荧光双向差示凝胶电泳技术操作过程

第二节　运用质谱进行定量的蛋白质组分析技术

正如前面所述，由于不同蛋白质和多肽在质谱仪中离子化能力不同，所以从质谱图中不能得到它们量的信息。但是，对于具有相同离子化能力的蛋白质或多肽，可以通过比较质谱峰的强弱，得到其相对量的数值。Oda 和 Gygi 等人利用此原理，用一种质量标签来标记一种状态下的蛋白质，然后与另一种状态下没有标记的蛋白质混合起来，进行质谱分析，然后通过比较某个蛋白质没有标记的峰和标记后峰的强弱，就可知道这种蛋白质在这种状态下表达量的变化。

一、稳定同位素代谢标记技术

该方法由 Oda[10]提出，其具体作法是，在两组酵母细胞平行生长的培养基中，一组含有天然氮同位素组分（^{14}N，99.6%；^{15}N，0.4%）；另一组基质相同，但富含 ^{15}N（>96%）。经过一段时间培养后，将细胞裂解，然后将这两组蛋白质等量混合，通过凝胶电泳分离和染色后，找出差异表达蛋白质，将其从凝胶中挖出，酶解后，用 MALDI-TOF 分析。由于 ^{15}N 标记的肽的相对分子质量要比 ^{14}N 标记的肽大，通过比较二者在质谱图上的强度，就可以精确定量同一蛋白质在不同生长条件下的表达量的变化。同时 Oda 等人运用这种方法准确的定量了蛋白质不同的磷酸化位点[10]。但是，这种方法也有它的缺点：①它只适用于对微生物的分析，而不能对从组织中提取的细胞直接进行分析；②同位素富集的基质环境可能会影响微生物的生长和蛋白质的生物合成；③同位素标记增加了被标记蛋白质的相对分子质量，从而加大了数据库搜寻的难度。

二、同位素亲和标签技术

该技术由 Gygi[7]提出，它主要是使用了一种称为 ICAT 化学试剂，其结构由三部分构成：第一，亲和标签（生物素，biotin），用来分离经 ICAT 标记后的多肽；第二，连接子，用来整合稳定的同位素；第三，活性基团，用来特异结合巯基（半胱氨

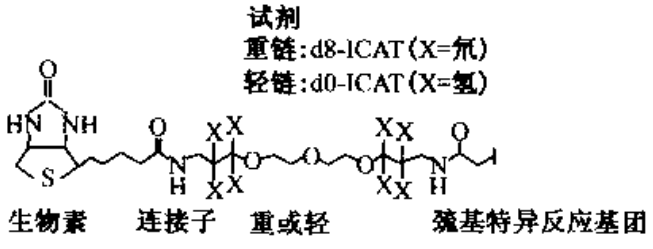


图 8-2　ICAT 试剂的结构图

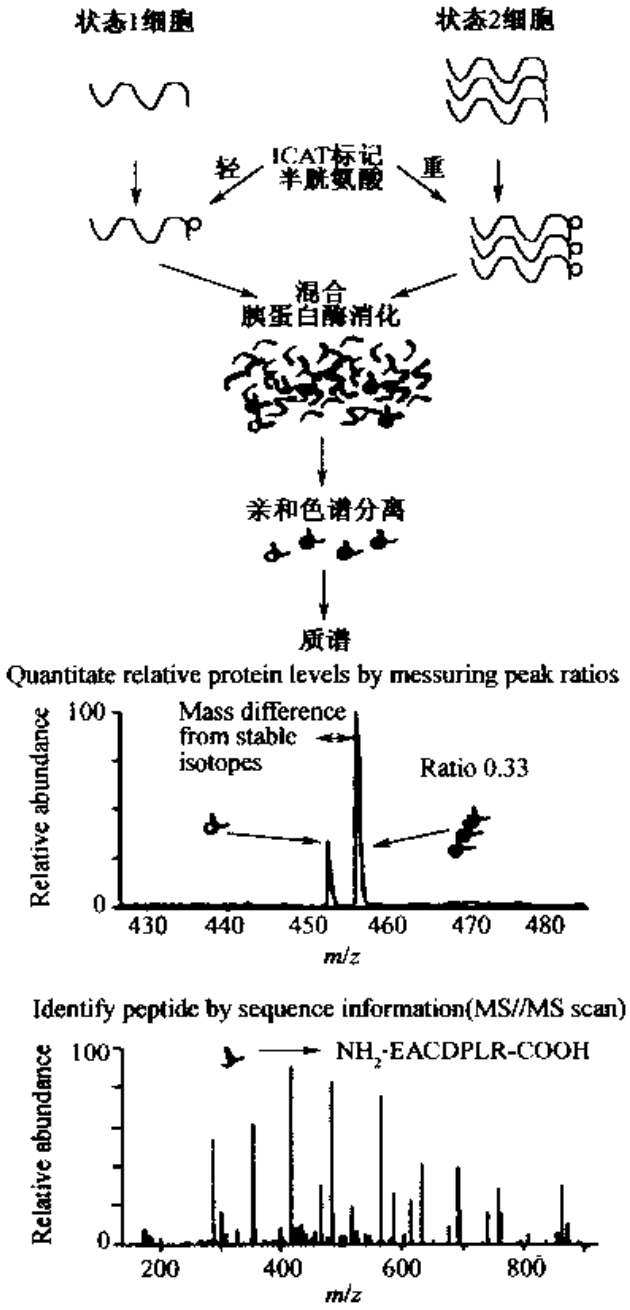


图 8-3　ICAT 战略图示

该图示显示了对整个细胞裂解物中单一蛋白的分析。分别收集两种不同状态下的细胞裂解物，变性，还原后，分别用“重”或“轻”ICAT 试剂来标记它们，等量混合后，蛋白酶解。运用生物素标签亲和色谱来分离被标记的多肽，通过在线的 HPLC 结合串联质谱对它们进行分析。ICAT 标记离子对的离子强度比例显示了在起始状态下其所属蛋白的相对含量。除此以外，串联质谱对肽序列的分析可以准确鉴定这个蛋白。这个战略可以定量和鉴定一个混合体系中所有的蛋白质，并且在理论上它可应用于像整个蛋白质组这样复杂的蛋白质混合体系。

酸)。试剂有两种形式,称之为"重"(连接子含有 8 个氘原子)和"轻"(连接子含有 8 个氢原子),如图 8-2 所示。含 8 个氢原子或 8 个氘原子的 ICAT 相对分子质量正好相差 8Da 或 4Da(带双电荷)。

当不同状态的细胞被裂解后,分别加入不同标记的 ICAT 与总蛋白质反应。ICAT 的反应基团会专一与蛋白质中的 Cys 共价结合,待充分反应后,将二者等量混合,再用胰蛋白酶酶解,HPLC 分离,就可对含生物素 ICAT 酶解片段进行串联质谱分析。如图 8-3 所示。

一对 ICAT 标记来源不同细胞状态下的相同肽段总是一前一后相邻分布在质谱图上,且相对分子质量正好相差 8Da 或 4Da(肽段带两个电荷)。通过 MS/MS 鉴定出该片段属于何种蛋白质后,再通过计算二者峰值的差异,就可知道这种蛋白在这两种细胞状态下表达的情况。

Gygi 等人对酵母乙醇脱氢酶的两种同工酶进行了分析。酵母的两种乙醇脱氢酶分别为 ADH1 和 ADH2。当酵母在有已糖的环境下生长时,ADH1 会将乙醛转化成乙醇。在无其他碳源存在时,乙醇会作为惟一的碳源被酵母利用,这一步依赖于 ADH2 将乙醇转化成醛。乙醛再经过其他反应进入代谢途径,以产生必需的碳源和能量供酵母的生长。ADH1 和 ADH2 的氨基酸同源性高达 93%,而且酶切后被分析的肽段只有一个氨基酸的差异(Val 到 Thr)。当半乳糖作为碳源时 ADH1 的表达量几乎是以乙醇为碳源的两倍。相反,ADH2 在乙醇诱导下的表达量比在半乳糖存在下高出 200 多倍,如图 8-4 所示。

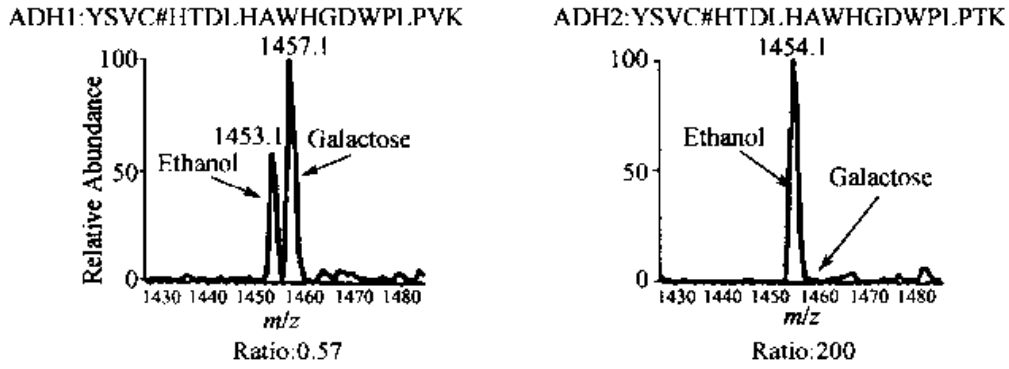


图 8-4 运用 ICAT 技术对在不同碳源环境中生长的酵母细胞中乙醇脱氢酶(ADH1 和 ADH2)的差异表达分析

ICAT 具有广泛的兼容性,主要表现在:第一,能够兼容分析任何条件下体液、细胞、细胞、组织中绝大部分蛋白质;第二,烷化反应即使在盐、去垢剂、稳定剂(如 SDS、尿素、盐酸胍等)存在下都可进行;第三,只需分析含 Cys 残基的肽段,从而降低了蛋白混合物分析的复杂性;第四,ICAT 战略允许任何类型的生化、免疫、物理的分离方法,因此能很好地定量分析微量蛋白质。当然,ICAT 技术依然存在一些

问题。首先，ICAT 的相对分子质量约为 500Da，这对肽段来说已经是一个很大的修饰物，这会增加数据库搜索算法的复杂性，对一些小的肽段（小于 7 个氨基酸）更是如此。其次，这一方法无法分析不含 Cys 的蛋白。在酵母中，约 8% 的蛋白质是不含有 Cys 残基的，但这一点完全可以通过合成对其他氨基酸基团专一的 ICAT 试剂对这部分蛋白质进行分析，这一思想将会有力地推动对蛋白质翻译后修饰的研究[6]。

最近，Griffin 等[14]运用 MALDI QqTOF 质谱和 ICAT 技术定量分析了以乙醇为碳源和以半乳糖为碳源情况下，酵母细胞的蛋白质组变化。酵母细胞在两种不同碳源存在下，生长达到稳定状态后，收集细胞，裂解后，总蛋白与 ICAT 试剂反应，然后用胰蛋白酶消化，将消化产物先过强阳离子交换树脂柱，再用反相毛细管液相色谱柱分离，最后将其馏分直接点到 MALDI 靶上，加上基质，就可对其进行质谱分析。他们用反相毛细管液相色谱柱分离 3 种不同阳离子交换树脂馏分，共点了 33 个靶点，分析得到了 59 种蛋白质，并对它们进行了定量分析。此外，用这种方法还鉴定了 8 个低拷贝蛋白质（密码子偏性系数小于 0.2），其中一个蛋白质的密码子偏性系数为 0.069。而 Smith 等人[15]用 ICAT 技术和傅里叶转换离子回旋共振质谱（fourier transform ion cyclotron resonance mass spectrometry，FTICR MS）相结合，得到了一种新的定量蛋白质组学研究技术。该技术主要运用了 FTICR 质谱超强的质量分辨能力，对细胞蛋白混合物进行一维分离后，就可进行鉴定。同上述一样的道理，与 ICAT 技术结合后，就可用于定量蛋白组分析。

第三节　蛋白质芯片技术

蛋白质芯片技术就是将一系列“诱饵”蛋白（如抗体）以阵列方式固定在经过特殊处理的底板上，然后将其与待分析的样品杂交，只有那些与“诱饵”结合的蛋白才被保留在芯片上。这种方法实质上是大规模的酶联免疫吸附测定，而酶联免疫吸附早已被应用于临床诊断。蛋白质芯片技术有两个关键因素：①空间上固定能辨别单一蛋白成分的分子。最明显的能够有效识别蛋白的分子是抗体，这种分子识别方式已被广泛应用于许多方面。已使用的抗体有单抗和多抗。如果用抗体作“蛋白质芯片”，必须遵循以下几个重要原则：首先，必须有支持获得和选择单个蛋白的方法，这个蛋白甚至是经翻译后修饰的。这个抗体必须是高度特异的，这意味着它们能够在一个复杂混合物中识别和区分单个蛋白。其次，混合物中单个蛋白与它们的抗体结合必须在相似的条件下进行，在一定程度上可以通过仔细控制抗体产生过程中筛选步骤（screening step）来达到。当然，抗体并不是惟一地用在“蛋白质芯片”上的蛋白质识别分子。任何分子基团只要它具有相应的亲合力并能特定辨认出单个蛋白质分子，都可以固定起来，形成“蛋白质芯片”；②检测混合物中单个蛋白与它们相应的识别分子间相互作用的方法。快速、平行和灵敏地检测蛋

白质与"蛋白质芯片"上认知分子间相互作用的方法，是对人们一个重大的挑战。现在发展的许多方法非常有效，并且也不需要复杂的装置，但是不适用于"蛋白质芯片"，如用第二种荧光标记认知分子检测俘获蛋白的方法[16]。蛋白质芯片技术的一种方式是将蛋白质芯片与已经经过不同的荧光标记物标记处于不同细胞状态下的细胞溶解物杂交，杂交后颜色及强度变化显现出结合在抗体上蛋白质含量的变化。这种方法必须明确知道抗体的特异性及其具体特征[17]。现已发展出三种不同类型的芯片包括玻璃板芯片、3D 胶芯片(3D gel pad chip)、微孔芯片(microwell chip)用作该技术的研究[18]。蛋白质芯片技术对分析蛋白质组的变化具有很大的应用潜力，但还有一系列问题需要解决，例如如何迅速获得大量可以用作"诱饵"的蛋白质；由于许多细胞膜蛋白只有在脂类存在情况下才具有活性，这些蛋白在提取过程中容易变性，对它们如何分析等[18,19]。Ciphergen 应用生物系统公司推出了一种基于蛋白质芯片——ProtenChip® 和质谱技术的蛋白质谱分析平台，其核心是一系列称为 ProtenChip® 的蛋白质芯片，它们具有多种特性，如阳离子交换、阴离子交换、金属离子亲和等。从细胞和体液中提取出的蛋白质混合物与一系列不同特性的 ProtenChip® 芯片相结合，洗去不与芯片结合的蛋白，然后用飞行时间质谱读出与芯片结合的蛋白，就得到了基于不同相互作用的多维蛋白质图谱。这种技术称为表面增强激光解析离子化(suface-enhanced laser desorption/ionization, SELDI)。与基质辅助激光解析电离技术(matrix-assisted laser desorption/ionization, MALDI)相比，它能够获得可重复的、统一的质谱图，因而可用于定量蛋白质研究，而 MALDI 却不行[20]。由于 ProtenChip® 不需要获得大量固定在芯片上的蛋白识别分子，因而简单、易行，已有许多成功的报道，具体的内容见参考文献 19。

随着蛋白质组学研究的不断深入和发展，定量蛋白质组学已成为当前研究的热点。目前定量蛋白质组研究的一个主要内容是研究在不同细胞状态下蛋白质的差异表达，特别是一些低拷贝蛋白质(如一些激酶和信号转导分子)。在这种情况下，研究者不仅要注重在不同细胞状态下蛋白质表达量的相对变化，还要注意其所用技术能定量的范围界限。从上面的讨论可以看出，人们已取得了一定的成功。但是，我们还应看到其中的不足。首先，所有的这些定量技术都只是一个相对的定量，还没有一个可以获得细胞或组织中蛋白表达的绝对的准确量的技术；其次，目前所用的这些技术也不是很完善，都还存在这样或那样的一些问题，如荧光染色技术相比银染技术而言，它的动力学线性范围有了很大的提高，但是它还不能符合细胞内蛋白质表达的动力学范围的要求；ICAT 技术还不能分析不含有 Cys 的蛋白质；人们还不能够获得用于蛋白质芯片技术的大量的蛋白质识别分子等。

参考文献

1 O'Farrell P H. High resolution two-dimensional electrophoresis of proteins. J Bio Chem, 1975, 250: 4007－

4021

2 Wasinger V C, Cordwell S J, Cerpa-Poljak A, Yan J X, Gooley A A, Wilkins M R, Duncan M W, Harris R, Williams K L, Humphery-Smith I. Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis, 1995, 16(7): 1090～1094

3 Opitech G J, Lewis K C, Jorgenson J W. Comprehensive on-line LC/LC/MS of proteins. Anal Chem, 1997, 69: 1518～1524

4 Rabilloud T, Strub J M, Luche S, van Dorssclaer A, Lunardi J. A comparison between Sypro Ruby and ruthenium II tris(bathophenanthroline disulfonate) as fluorescent stains for protein detection in gels. Proteomics, 2001, 1: 699～704

5 Rabilloud T. Detecting proteins separated by 2-D gel electrophoresis. Anal Chem, 2000, 72(1): 48A～55A

6 Mann M. Quantitative proteomics? Nat biotechnol, 1999, 17: 954～955

7 Gygi S P, Rist B, Gerber S A, Turecek F, Gelb M H, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol, 1999, 17: 994～999

8 Blackstock W P. Proteomics: quantitative and physical mapping of cellular proteins. Trends Biotechnol, 1999, 17: 121～127

9 Gygi S P, Rist B, Aebersold R. Measuring gene expression by quantitative proteome analysis. Curr Opin Biotech, 2000, 11: 396～401

10 Oda Y, Huang K, Cross F R, Cowburn D, Chait B T. Accurate quantitative of protein expression and site-specific phosphorylation. Proc Natl Acad Sci USA, 1999, 96: 6591～6596

11 Rabilloud T. Dectecting Proteins. Anal chem, 2000, 72(1): 48A～54A

12 Lopez M F, Berggren K, Chernokalskaya E, Lazarev A, Robinson M, Patton W F. A comparison of silver stain and SYPRO Ruby Protin Gel stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. Electrophoresis, 2000, 21: 3673～3683

13 Tonge R, Shaw J, Middleton B, Rowlinson R, Rayner S, Young J, Pognan F, Hawkins E, Currie I, Davison M. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. Proteomics, 2001, 1(3): 377～396

14 Griffin T J, Gygi S P, Rist B, Aebersold R, Loboda A, Jilkine A, Ens W, Standing K G. Quantitative proteomics analysis using a MALDI quadruple time-of-flight mass spectrometer. Anal chem, 2001, 73(5): 978～986

15 Smith R D, Pasa-Tolić L, Lipton M S, Jensen P K, Anderson G A, Shen Y, Conrads T P, Udseth H R, Harkewicz R, Belov M E, Masselon C, Veenstra T D. Rapid quantitative measurements of proteomes by Fourier transform ion cyclotron resonance mass spectrometry. Electrophoresis, 2001, 22: 1652～1668.

16 Pennington S, Dunn M J. Proteomics: From Protein Sequence to Function. New York: Springer-Verlag New York Inc, 2001, 214～220

17 Pandey A, Mann M. Proteomics to study genes and genomes. Nature, 2000, 405(6788): 837

18 Zhu H, Snyder M. Protein arrays and microarrays. Curr Opin Chem. Biol, 2001, 5(1): 40～45

19 杨何义，应万涛，钱小红. 蛋白质组学技术研究进展. 自然科学进展，2002, 12(1): 13～17

20 Fung E T, Thulasiraman V, Weinberger S R. Protein biochips for differential profiling. Curr Opin Biotechnol, 2001, 12: 65～69

（杨何义）

第九章　蛋白质组研究中的生物信息学

第一节　生物信息学简介

一、什么是生物信息学

生物信息学是随着人类基因组计划(Human Genome Project, HGP)而发展起来的。美国科学家于1985年率先提出人类基因组计划,其科学目标是绘制遗传图和物理图,旨在阐明人类基因组30亿个碱基对的序列,发现所有人类基因并搞清其在染色体上的位置,破译人类全部遗传信息,使人类第一次在分子水平上全面地认识自我。计划于1990年正式启动,目前,有美国、德国、日本、英国、法国和中国6个国家的科学家正式加入了这一计划。2001年2月,*Science*(2001年2月16日)和 *Nature*(2001年2月15日)分别刊登了由私营公司塞莱拉和政府资助的人类基因组计划绘制的基因组草图,精确的人体基因组图谱有望于2003年前最终绘制出来。

20世纪生物学最宏伟的"人类基因组计划"的顺利实施,大大加速了生命科学各方面的发展。21世纪初,人和其他模式生物(微生物、线虫、果蝇、斑马鱼、拟南芥、水稻等)基因组作图和测序将陆续完成。分子生物学研究的重点也将从基因组扩展到蛋白质组。在这种发展形势下,生物学正进入"后基因组时代"(postgenome era),或者说功能基因组和蛋白质组时代。从总体看来,以基因组研究为核心,在方法学上分析与综合相结合,比较和实验相结合,微观与宏观相结合,来探讨生命的本质和起源,遗传、发育和进化的理论大综合,以及阐明脑高级功能活动,将是21世纪生物学基础理论研究的大趋势[1]。"人类基因组计划"和"曼哈顿工程"和"阿波罗登月计划"并称20世纪的三大科学计划。

生物信息学(bioinformatics)是一门新兴的交叉学科。它包含了生物信息的获取、处理、存储、发布、分析和解释等在内的所有方面,它综合运用数学、计算机科学和生物学的各种工具,来阐明和理解大量数据所包含的生物学意义。具体地说,生物信息学是把基因组DNA序列信息分析作为源头,找到基因组序列中代表蛋白质和RNA基因的编码区;同时,阐明基因组中大量存在的非编码区的信息实质,破译隐藏在DNA序列中的遗传语言规律;在此基础上,归纳、整理与基因组遗传信息释放及其调控相关的转录谱和蛋白质谱的数据,从而认识代谢、发育、分化、进化的规律。生物信息学还利用基因组中编码区的信息进行蛋白质空间结构的模拟

和蛋白质功能的预测，并将此类信息与生物体和生命过程的生理生化信息相结合，阐明其分子机理，最终进行蛋白质、核酸的分子设计、药物设计和个体化的医疗保健设计[2]。理解大量生物学数据所包括的生物学意义已成为后基因组时代极其重要的课题。生物信息学的作用将日益重要。有理由认为，今日生物学数据的巨大积累将导致重大生物学规律的发现。

"人类基因组计划"在完成基因组全部序列测序后，下一步更艰巨的任务是读懂基因组的工作语言——遗传语言破译。这是21世纪自然科学面临的最大挑战之一。"人类基因组计划"的完成和深入发展，将有可能从更深层次上了解人体生长、发育、正常生理活动和各种疾病的病因及发病机理，并提出防治策略、途径和方法。生物信息学的重点发展方向是[1]：

1） 人类基因组信息结构复杂性；序列(特别是非编码区)信息分析；

2） 基因组结构与遗传语言：语法和词法分析；

3） 大规模基因表达谱分析，相关算法、软件研究；基因表达调控网络研究；

4） 基因组信息相关的蛋白质功能分析；

5） 生物信息学中新理论、新方法、新技术和新软件研究。

在当前基因组信息爆发的时代，建立超大规模计算系统，发展全新的生物信息学的理论、方法，分析这些数据，从中获得生物体结构、功能的相关信息基因组研究取得成果的决定性步骤。生物学是生物信息学的核心和灵魂，数学与计算机技术则是它的基本工具。

随着破译生命密码的人类基因组计划进入尾声，一个以蛋白质和药物基因学为研究重点的后基因组时代已经拉开序幕，并将提出更为严峻的挑战。美国塞莱拉公司首席科学家文特尔在科学年会的基调演讲中说，人类基因组、蛋白组和药物是生命科学研究路上的三个阶段，塞莱拉公司现已进入第二阶段，即蛋白组研究。塞莱拉为此增添了大批蛋白质鉴别和分析设备，目的是每天对数百万个蛋白质片段进行识别和分类，最终绘制出一张蛋白组图[3]。

生物信息学是当今生命科学和自然科学的重大前沿领域之一，同时也将是21世纪自然科学的核心领域之一。

二、生物信息学的研究内容

生物信息学具有以下研究内容[4]。

1. 序列比对

基本问题是比较两个或两个以上符号序列的相似性或不相似性。序列比对(alignment)是生物信息学的基础，非常重要。两个序列的比对有较成熟的动态规划算法，以及在此基础上编写的比对软件包——BALST和FASTA，可以免费下载

使用。这些软件在数据库查询和搜索中有重要的应用。有时两个序列总体并不很相似,但某些局部片段相似性很高。Smith-Waterman 算法是解决局部比对的好算法,缺点是速度较慢。两个以上序列的多重序列比对目前还缺乏快速而又十分有效的算法。

2. 结构比对

基本问题是比较两个或两个以上蛋白质分子空间结构的相似性或不相似性。已有一些算法。

3. 蛋白质结构预测,包括二级和三级结构预测,是最重要的课题之一

从方法上来看有演绎法和归纳法两种途径。前者主要是从一些基本原理或假设出发来预测和研究蛋白质的结构和折叠过程。分子力学和分子动力学属这一范畴。后者主要是从观察和总结已知结构的蛋白质结构规律出发来预测未知蛋白质的结构。同源模建和指认(threading)方法属于这一范畴。虽然经过 30 余年的努力,蛋白结构预测研究现状远远不能满足实际需要。

4. 计算机辅助基因识别(仅指蛋白质编码基因)

基本问题是给定基因组序列后,正确识别基因的范围和在基因组序列中的精确位置。这是最重要的课题之一,而且越来越重要。经过 20 余年的努力,提出了数十种算法,有十种左右重要的算法和相应软件上网提供免费服务。原核生物计算机辅助基因识别相对容易些,结果好一些。从具有较多内含子的真核生物基因组序列中正确识别出起始密码子、剪切位点和终止密码子,是个相当困难的问题,研究现状不能令人满意,仍有大量的工作要做。

5. 非编码区分析和 DNA 语言研究,是最重要的课题之一

在人类基因组中,编码部分占总序列的 3%～5%,其他通常称为"垃圾"DNA,其实一点也不是垃圾,只是我们暂时还不知道其重要的功能。分析非编码区 DNA 序列需要大胆的想象和崭新的研究思路和方法。DNA 序列作为一种遗传语言,不仅体现在编码序列之中,而且隐含在非编码序列之中。

6. 分子进化和比较基因组学,是最重要的课题之一

早期的工作主要是利用不同物种中同一种基因序列的异同来研究生物的进化,构建进化树。既可以用 DNA 序列也可以用其编码的氨基酸序列来做,甚至可通过相关蛋白质的结构比对来研究分子进化。以上研究已经积累了大量的工作。近年来由于较多模式生物基因组测序任务的完成,为从整个基因组的角度来研究分子进化提供了条件。可以设想,比较两个或多个完整基因组这一工作需要新的

思路和方法，当然也渴望得到更丰硕的成果。这方面可做的工作是很多的。

7. 序列重叠群装配

一般来说，根据现行的测序技术，每次反应只能测出 500bp 或更多一些碱基对的序列，这就有一个由大量的较短的序列全体构成的重叠群(contigs)。逐步把它们拼接起来形成序列更长的重叠群，直至得到完整序列的过程称为重叠群装配。拼接 EST 数据以发现全长新基因也有类似的问题。已经证明，这是一个 NP 完备性算法问题。

8. 遗传密码的起源

遗传密码为什么是现在这样的？这一直是一个谜。一种最简单的理论认为，密码子与氨基酸之间的关系是生物进化历史上一次偶然的事件而造成的，并被固定在现代生物最后的共同祖先里，一直延续至今。不同于这种"冻结"理论，有人曾分别提出过选择优化、化学和历史等三种学说来解释遗传密码。随着各种生物基因组测序任务的完成，为研究遗传密码的起源和检验上述理论的真伪提供了新的素材。

9. 基于结构的药物设计

人类基因组计划的目的之一在于阐明人的约 10 万种蛋白质的结构、功能、相互作用以及与各种人类疾病之间的关系，寻求各种治疗和预防方法，包括药物治疗。基于生物大分子结构的药物设计是生物信息学中的极为重要的研究领域。为了抑制某些酶或蛋白质的活性，在已知其三级结构的基础上，可以利用分子对接算法，在计算机上设计抑制剂分子，作为候选药物。这种发现新药物的方法有强大的生命力，也有着巨大的经济效益。

10. 生物信息处理并行算法的研究

由于生物信息数据的规模极其巨大，因此国内外都开展了生物信息处理算法并行化方向的研究。国外还开展了特殊生物信息处理中算法的研究以及在硬件基础上的并行化方向的研究，主要是研究生物信息学中的一些关键的算法，研究其中的可并行性，然后将其固化到硬件芯片中，从而提高整个计算系统的性能。

11. 其他

如基因表达谱分析，代谢网络分析，基因芯片设计和蛋白质组学数据分析等，逐渐成为生物信息学中新兴的重要研究领域。

三、我国生物信息学研究的目标

2001年9月，国家科学技术部对“生物和现代农业技术领域生物信息技术主题”提出的目标是[5]：

实现生物技术与信息技术以及其他学科的结合，实现基因组数据、蛋白质组和结构基因组数据、天然及合成化合物数据的计算机处理、分析和可视化，以及生物实验和生物分子的模拟设计，解析蛋白质三维结构和蛋白质组的时空表达关系等，提高生物信息处理、分析和利用的水平，为我国生命科学和生物技术的源头创新奠定基础。

主要内容是：

1. 生物信息的获取与开发：建立国家级的生物信息搜集、分析和服务基地，形成开放式和标准化的生物信息数据库应用、服务体系和技术平台，为我国生物技术的研究提供高效服务；重点建成一批有特色、可整合的一级和二级数据库。

2. 生物信息加工和利用：建立用于基因组、功能基因组和高通量生物技术研究，可整合各种相关信息及分析计算方法、界面友好的数据分析技术平台，开发出若干具有自主知识产权、有商业化前景的计算机分析系统和生物信息学软件包。

3. 结构基因组和蛋白质组学研究：建立有一定竞争力的结构基因组学研究和蛋白质组学研究的技术平台，完成一批新蛋白质或有重要价值的蛋白质及其复合物的结构测定；初步完成蛋白质组信息处理、分析和解释的生物信息学方法与工具的研制，提供对重大疾病的预警、诊断和治疗有应用前景的候选靶标。

4. 高通量药物筛选及相关技术：建立高通量药物筛选新模型；建立具有结构多样性的化合物样品库；完善高通量药物筛选和组合化学新技术平台；发现一批具有自主知识产权的创新药物先导化合物。

5. 小分子药物设计和分子设计：发现具有新机理、新结构类型的原创性先导化合物，获得一批国际专利；针对药物设计及分子设计的关键环节，研发具有自主知识产权可供推广应用的软件并获得版权；获得应用前景明确的蛋白质及其他生物功能分子。

6. 生物芯片：建立和完善生物芯片的技术平台，研制有重要应用价值的基因组表达芯片及检测专用生物芯片，开发出实用的生物技术和基因组研究、医学诊断、农产品开发等生物芯片新技术、新方法、新产品。

7. 化学创新药物与新剂型：创新药物研究需在三年内完成临床前阶段的综合评价，确证其是否具有临床应用前景；新剂型研究需针对重要疾病，研发出市场覆盖面广、需求量大、具有自主知识产权的新制剂或新辅料，在三年内可以获得临床批文或新药证书。

四、蛋白质组研究中生物信息学的任务

1. 蛋白质组学研究的目的和任务

20 世纪中期以来，随着 DNA 双螺旋结构的提出和蛋白质空间结构的 X-射线解析，开始了分子生物学时代，对遗传信息载体 DNA 和生命功能的主要体现者蛋白质的研究，成为生命科学研究的主要内容。20 世纪 90 年代初期，美国生物学家提出并实施了人类基因组计划，预计用 15 年的时间，30 亿美元的资助，对人类基因组的全部 DNA 序列进行测定，希望在分子水平上破译人类所有的遗传信息，即测定大约 30 亿碱基对的 DNA 序列和识别其中所有的基因（基因组中转录表达的功能单位）。经过各国科学家多年的努力，人类基因组计划已经取得了巨大的成绩，一些低等生物的 DNA 全序列已被阐明，人类 DNA 序列也基本已测定，迄今已测定的表达序列标志（EST）已大体涵盖人类的所有基因。在这样的形势下，科学家们认为，生命科学已经进入了后基因组时代。

在后基因组时代，生物学家们的研究重心已经从解释生命的所有遗传信息转移到在整体水平上对生物功能的研究。这种转向的第一个标志就是产生了一门成为功能基因组学（functional genomics）的新学科。它采用一些新的技术，如 SAGE、DNA 芯片，对成千上万的基因表达进行分析和比较，力图从基因组整体水平上对基因的活动规律进行阐述。但是，由于生物功能的主要体现者是蛋白质，而蛋白质有其自身特有的活动规律，仅仅从基因的角度来研究是远远不够的。例如蛋白质的修饰加工、转运定位、结构变化、蛋白质与蛋白质的相互作用、蛋白质与其他生物分子的相互作用等活动，均无法在基因组水平上获知。

正是因为基因组学（genomics）有这样的局限性，于 20 世纪 90 年代中期，在人类基因组计划研究发展及功能基因组学的基础上，国际上萌发产生了一门在整体水平上研究细胞内蛋白质的组成及其活动规律的新兴学科——蛋白质组学（proteomics），它以蛋白质组（proteome）为研究对象。蛋白质组是指“由一个细胞或一个组织的基因组所表达的全部相应的蛋白质”。测定一个有机体的基因组所表达的全部蛋白质的设想，萌发在 1975 年双向凝胶电泳发明之时。1994 年 Williams 正式提出了这个问题，而“蛋白质组”的名词则是由 Wilkins 创造的，发表在 1995 年 7 月的 *Electrophoresis* 杂志上。

蛋白质组与基因组相对应，但二者又有根本不同之处：一个有机体只有一个确定的基因组，组成该有机体的所有不同细胞共同享用一个确定的基因组；而蛋白质组则是一个动态的概念，它不仅在同一个机体的不同组织和细胞中不同，在同一机体的不同发育阶段，在不同的生理状态下，乃至在不同的外界环境下都是不同的。正是这种复杂的基因表达模式，表现了各种复杂的生命活动，每一种生命运动形式，都是特定蛋白质群体在不同时间和空间出现，并发挥功能的不同组合的结果。

基因 DNA 的序列并不能提供这些信息，再加上由于基因剪接、蛋白质翻译后修饰和蛋白质剪接，基因遗传信息的表现规律就更加复杂，不再是经典的一个基因一个蛋白的对应关系，一个基因可以表达的蛋白质数目可能远大于一。对细菌，可能为 1.2～1.3；对酵母则为 3；而对人，可高达 10。后基因组和蛋白质组研究，是为阐明生命活动本质所不可缺少的基因组研究的更为复杂的后续部分，无疑将成为 21 世纪生命科学研究的主要任务。

2. 蛋白质组研究中生物信息学要解决的问题

蛋白质组学的第一个出发点是深入研究细胞的活性成分——蛋白质的动态信息，第二个出发点是假定任何细胞或组织的全部蛋白质在本质上是相关的，研究它们之间的相互关系对于理解其生活史与进化特点也是极为重要的。大规模蛋白质组分析过程主要包括：样品制备、图像分析、蛋白质成分的分析与鉴定等[6]。

人类蛋白质组组织(the Human Proteome Organization, HUPO)对生物信息学提出的目标是：开发数据库、分析软件和注释标准。

我们现有哪些生物信息学基础解构元素，还需要建立或改进哪些？

生物信息学基础解构元素包括：现有的和预期的数据类型、数据提交、数据库互操作性、注释标准、数据库/服务及数据收集的资助，等等。

现有的和预期的数据类型包括序列数据、蛋白质功能注释、解构数据、阵列数据、2D-PAGE、色谱数据、质谱数据、PPI 数据，等等。

数据提交：如何鼓励/加强数据提交？出版社和基金会又在其中扮演什么样的角色？

数据库互操作性：如何选择分布式与集中式数据库？它们分别有哪些特点？

注释标准：约束词汇表、证明属性。

在当前急切要制定以下的蛋白质组学标准：数据和注释标准、数据格式标准、质谱和 PPI 数据的最初应用、数据产出者、数据库和软件生产者，等等。

第二节　数据库的构建

数据库是生物信息学的主要内容，各种数据库几乎覆盖了生命科学的各个领域。手工搜集的蛋白质数据库早在 20 世纪 60 年代就在美国开始建立。美国洛斯阿拉莫斯国家实验室 1979 年开通的基因库 GeneBank，现在由 1988 年成立的国家生物信息中心(NCBI)管理维护。欧洲分子生物学实验室 1982 年开始服务的 EMBL 数据库和随后建立的欧洲生物网(EMBNet)，1994 年改由当年建在英国剑桥的欧洲生物信息研究所(EBI)管理。日本 1984 年着手建立国家级的核酸数据库 DDBJ，1987 年正式服务。目前绝大部分核酸和蛋白质数据由美国、欧洲和日本产生，以上三家共同组成了 DDBJ/EMBI/GeneBank 国际核酸序列数据库，每天交

换数据，同步更新。其他国家如德国、法国、意大利、澳大利亚、丹麦、以色列等，在分享网络资源的同时，还纷纷建立自己的生物信息中心，为本国服务[7]。

生物信息学数据库覆盖面广，分布分散且格式不统一，因此一些生物计算中心将多个数据库整合在一起提供综合服务，如 EBI 的 SRS(Sequence Retrieval System)包含了核酸序列库、蛋白质序列库、三维结构库等 30 多个数据库及 CLUSTALW、PROSITESEARCH 等强有力的搜索工具，用户可以进行多个数据库的多种查询。

从细菌到人类，众多物种的基因和蛋白质数据正在以科学史上从未有过的高速度增长。生物学数据的积累并不仅仅表现在 DNA 序列方面，与其同步的还有蛋白质一级结构，即氨基酸序列的增长。此外，迄今为止，已有一万多种蛋白质的空间结构以不同的分辨精度被测定，而这种增长速度本身也在加速。随着生物信息数据的爆炸性增长，各种数据库的建立和管理，成为所有生物信息学工作的基础。

一、工程前期工作

1. 用户调研

用户调研的目的是收集需求，其方法是直接面对面与用户交流，最好能亲身体验用户工作开展的过程，注重了解工作的细节。用户调研的结束标志是完成了需求分析设计文档，并且得到用户的认可。在工程前期开展好用户调研对于工程的顺利进行和完成非常重要。

用户调研的人员应当集中、固定，具有多方面的经验，如用户交流经验、程序设计经验、系统分析经验等，但调研人员数量不能太多，并能对重大问题进行独立决策。

2. 需求分析

需求分析的目的是明确用户需要实现的内容。

可以采取多样的需求分析方法和技术，但是根本上必须描述清楚用户的业务工作、软件需求。有很多的需求分析的标准和规范可以采用，也可以自己根据实际情况制定需求分析的规范。需求分析的人员可与调研人员重合，但要增加软件分析人员，并为调研人员提出调研的内容和方向。需求分析结束的标志是程序员能够开始进行程序设计，并和建立系统原型的工作结合起来进行，因此只有程序员才能评判需求调研的质量。

需求分析阶段投入的时间一定不能少于全部工程时间的 30%，当然 40% 会更好，否则很难保证需求分析的深度和质量。在需求分析阶段至少应建立三类组织：调研组、需求分析组、原型开发组(可分成多个小组)，统一由项目负责人负责。

3. 建立系统原型

建立系统原型是完成需求分析的补充手段，目的是明确和确定用户需求，引导用户提出更具体的内容。建立系统原型要求使用适当的工具和容易的软件环境，满足快速开发的要求。

系统原型一般不作为将来软件系统实现的基础，因为此时的原型系统只考虑反映用户需求的情况，缺乏统一的考虑，因此最好在设计、实现阶段重新开始。应避免使用户产生“软件开发已经开始，并会很快完成”的观点，一定要明确此时软件还在前期的分析阶段，距离设计、实现还有很长的距离。

4. 建立总体实施方案

建立系统的总体实施方案是必需的，目的是使甲乙双方共同认可工程的实施计划，保持一定的耐心和总体意识。总体实施方案应在需求分析基本完成的基础上开始。

总体实施方案应说明：任务量、计划进度、资金、实施阶段、各阶段结束的标志和开始的条件、完成时提交的内容。总体建设方案应在项目负责人、甲方负责人、乙方负责人共同认可的情况下，投入实施指导，否则要继续调整，此时项目负责人一定要做好乙方负责人的工作，争取更多的时间。

一旦总体实施方案确定，就必须按照实施方案的阶段安排逐步开展工作，并进行量化和考核，如果某一个阶段没有完成，则必须对照检讨，采取措施。否则将造成项目实施组织的混乱，因为大家无法再按照一致的实施方案开展工作了，为工程项目的顺利开展提出了警告。

5. 建立组织机构、制度

一定要明确组织机构是必不可少的，不要回避和搁置这个问题。只有建立了工程组织机构和管理制度，才具备完成项目的基础，因为工作的完成归根结底是人的协作。由于大型项目需要的人员比较多，因此没有完善的组织机构和制度是无法协调好各个部分的工作的。

组织机构的建立应在项目开始需求调研的时候就开始着手建立，在众多制度中“个人激励制度”是影响最大的一个制度。

二、软件系统实现

1. 系统设计

系统体系结构的确定是为软件需求服务的，没有必要为了结构而结构。在不同情况下采取不同的体系结构是明智的。体系结构的确定受到软件用户数量、运

行环境、更新维护要求、软件分布等外部因素的影响。

目前有很多系统设计的方法，选择某一方法(OMT、UML、OO)要根据系统的特点来确定。针对信息系统而言，可以自己制定合适的分析设计方法，因为目前虽然有很多的方法，但不太适合信息系统建设的要求。在信息系统设计的时候要灵活应用各种分析方法，照搬哪一种方法都不太合适。

在实际工作中，大家都知道文档很重要，但是经常和编程工作产生矛盾，不能得到及时的更新。为此在项目进行过程中，需要确定当前工程文档的各种标准，以及工程需要哪几类文档，文档提交和发布的时机等内容。这些文档包括：工作规范文档、分析设计文档、工作计划文档、日志文档、任务分配考核文档等。

在软件开发工程中，不是采取越多新技术越好，软件的设计也不要走入"尽量提高水平"的误区。由于负责软件设计的技术人员具有追求完美的特点，主观上尽力想将系统设计得完美，因此会造成软件实现需要的工作量增大，同时用户又不一定接受这些内容。在考虑软件水平的时候，主要考虑新技术的采用和成熟技术的采用要成比例。

系统的设计起主导作用之一的因素是用户的特点。不论什么特点的用户都采用一致的设计思路、风格、特点不是太好的选择，脱离对用户实际情况的考虑的设计太学术化，并且在设计的时候要照顾到系统的可维护性、易用性、实现的容易性等问题。其中原则之一是：能简化设计就简化，降低设计的复杂度。不要采用太复杂的设计，因为这样的设计不便于理解和维护。

2. 程序质量

程序质量是项目全方位的体现，包括管理水平、程序员的工作状况、相应的标准规范、文档的考核、个人激励制度等等所有的因素，这里仅就狭义的范围，讨论一下程序质量。

多人协作编写一个工程项目的工作的确具有挑战性，要完成好项目工作，没有统一的标准规范是不可能的。标准规范是为程序设计服务的，因此要尽量消化，不要模糊照搬，那样不会有帮助。程序设计要制定的规范包括：如何注释、注释量、变量命名规范、函数程序命名规范等，并针对这些规范的要求详细考核每个程序的设计情况。

编码质量包括：正确、易读、易维护。正确取决于设计文档和程序员的水平，并且可以通过控制编码的易读和易维护来保证编码的正确性，并且编码的质量控制需要量化指标和规范。

在大型信息系统开发工程中，采用适当的版本控制工具对程序代码进行控制是非常必要的。

文档的管理主要是更新、维护，并且要持之以恒，坚持到底。在程序设计过程中以一定的力度要求，在不同阶段要完成不同阶段的文档，这样才能以文档作为协

同工作的基础。这点对于大型信息系统的开发非常重要。

编程工作在需求分析阶段、程序实现阶段、系统运行维护阶段等都要有时间的投入，认识到这一点的意义在于：为了保证每个阶段的编程时间投入得最经济，就要控制好程序的质量。假设程序设计阶段代码很潦草，那么在运行维护阶段就要受到惩罚。

纵观众多得到用户认可的信息系统界面的共同特点是简洁，从而易学易用。如果用户界面组织得过于复杂，对于众多计算机水平一般的用户而言就难以掌握。所以在系统实现的过程中，要不断地与用户交流，确定用户喜欢的界面风格。

3. 组织管理

除了技术本身之外，另一个影响项目进展和质量的重要因素是组织管理水平。这里有两个问题：组织水平和管理水平。一般"组织"、"管理"总是放在一起使用，实际上组织是管理发挥作用的基础。即使是一个人总负责，这也是一群人组成的组织，也要有组织制度。所以，在实施过程中，建立组织制度、提高管理水平也是信息系统工程建设需要考虑的问题。

就管理的客体而言，无非是人、财、物，并通过合理搭配完成一定的任务，这些问题有很多需要研究的内容。对人的管理包括：建立组织、建立制度、明确责任、权力、利益；对财而言，做好预算、合理分配、收支分明；对物而言做到合理分配、动态调整；对完成任务而言，做好规划、计划、监督考核等。

组织是管理施加的客体，对于人数相对较少的项目，管理可以针对每个人，而对于人数庞大的项目，管理就要施加到组织上去。这里所说的组织可以根据项目的不同进行不同的组织，最基本的组织之一就是"程序组"，主要职责是完成程序编码，一个程序组就是项目负责人管理施加的客体，而对于组员来讲则是组长施加的客体。所以可以看出没有合理和稳定的组织，项目的管理就难以实施。

管理的本质是解决问题，如果在项目过程中没有问题，那么管理就不需要存在，而实施恰恰相反，在项目开发过程中有很多的问题，因而管理工作不但繁重而且重要。最日常的工作在于计划的制定、任务的安排、工作考核等等。要完成好管理不仅是项目负责人一个人的事情，它需要组织内有机的联系自动地对管理产生反应，得以落实管理的思想。

工作计划有长期的和短期的，长期可以是整个项目实施期内的计划，短期可以是一天的计划，对于小组比较适用周计划和日计划，而对于项目组要详细制定月计划和周计划来控制项目的进度并作为工作的考核依据。

任何情况下都难免会有意外发生，为此在管理以及制定计划的时候一定要考虑弥补措施，做到防患于未然。这些措施包括：突发的技术难题、突然的人员更换、突然的用户需求、突然的计划变动等等。

好的项目开发要建立稳定的开发环境，并要确保项目的开发环境与目标环境

的一致性，否则将造成巨大的困难。保持环境的相对稳定也非常重要，即在项目的开发过程中环境不要变动，包括系统的升级、不必要的或相关的软件的更换等等。如果目标环境为 Oracle 8i，使用 Oracle 9i 来作为开发环境则会产生很多不必要的问题。

大型信息系统一般经历的时间比较长，参与的人员比较多，因此进行必要的培训和知识更新是非常必要的。尤其是在项目开始实施之前，针对项目的若干总体要求要详细的说明并力求达成共识，这为今后大家很好的协同工作创造了一个基础。培训的内容是多方面的：技术、标准、规范、管理制度等等。

4．实现技术

数据库的设计是信息系统设计的基础。设计数据库时，在考虑设计要求的前提下，尽量考虑如何简化系统的实现。因为有的时候好的数据库设计会大大增加系统实现的复杂度和工作量。这需要不断的积累经验。

多人参与的信息系统设计要保持各个程序开发语言版本的高度一致，这个问题很容易被忽略，但是到系统集成的时候会造成很多的问题。为此，最好所有的人都适用同一个安装程序。除了开发语言之外的其他软件环境组成的开发平台也要保持一致和稳定，才能减少系统集成时出现的问题。

新技术的采用是一把双刃剑，如果采用过多会阻碍系统的开发进度。因此要从实际考虑，采用合理数量的新技术。这一点还受程序员的接受能力等因素的影响，如果接受能力强，则不必太担心。

5．软件测试

程序员的测试在很多软件测试的文献中经常被忽略，如果程序员提交的程序代码，他自己从未进行过测试，就提交给测试人员测试是不可想象的。程序员测试的一个基本原则是要确保所有的代码在正确使用下不会有问题，否则将浪费大量的时间修改代码。

对于测试的问题大家都有一致的认识，但在项目实施中，关键是按照规范要求组织项目测试小组。这主要受到人员的限制，即组织有经验的测试人员跟踪测试是比较困难的，同时也要有一定的投入。但是不论如何，还是要建立一个对系统进行全面测试的测试小组，随着系统的初步实施进行系统测试。测试的工作量是巨大的，应得到正确的和足够的认识，否则交付的软件质量无法保证，那么项目是否成功就没有保证。应该认识到不仅按时完成重要，按质完成更重要。

面向应用的项目，在交付用户正式使用之前要经过一定时间的用户测试，否则就为今后不断地出现问题埋下了伏笔。因此，针对用户测试要组织好，同时还要对资金、人力投入做一个合理的预算。

三、实施及维护

一般认为项目到实施维护阶段，就基本上结束了。实际并非如此，因为就软件开发周期而言，实施阶段的时间投入还要占一定的比例，并且在这个阶段由于对系统的切身使用，多数用户会提出很多的意见和建议，抛开很多不合理问题，还有很多问题需要解决，这就要求项目组还不能放松，还要再紧张地工作一段时间。

1. 管理者完成的工作

项目实施过程中管理者要完成的工作是非常繁杂的，在此仅做简单的说明，更多的内容要在实践中多思考、多总结，不断地提高。

1）树立用户意见至上的意识，并让项目组全体普遍接受。

2）做好与用户的沟通，尤其针对不合理的问题，要给出合理的解释。

3）使用户欣然接受你交付的系统是不断努力才能有的结果。

4）要做到不急不躁，同时要注意调整程序员的情绪。

2. 关于维护阶段的投入

维护阶段除了项目设计阶段的投入外，还要增加投入，包括：现场维护的费用、培训的费用(尤其是组织培训班)、人员的增加(增加与用户交互的人员，再次启用调研组)等等。

3. 加强与用户的沟通

加强与用户的沟通不只是实施阶段应当注意的，但是实施阶段是使用户产生意见的最关键的阶段。首先作为项目的双方负责人要做到及时、有效的沟通；其次，各个程序组在与普通用户进行沟通的时候也要注意方式、方法，同时要尽量能用简洁的语言给出明确的解释和说明。需要说明的是：没有每一个用户的认可和支持就是失败。

四、MySQL 数据库

构建数据库首先要考虑的是选用的硬件和操作系统平台以及要使用的数据库管理系统(DBMS)。各种操作系统下都有大量可供选用的数据库管理系统。许多程序员习惯在 Microsoft Windows 平台下进行开发工作，该平台下有许多可视化的开发工具，但很少有大型数据库使用 Windows 操作系统，而且操作系统和数据库管理系统、开发工具等也需要一笔不小的费用。大型数据库一般在 Unix 系统下运行，其安全性、稳定性、性能和服务都能得到充分的保证。Linux 是一种免费的

类 Unix 操作系统，近年来发展很快，成为很多中小企业服务器操作系统的首选。许多商业数据库（如 IBM 的 DB2，Oracle，Informix，Sybase 等）也提供了 Linux 版本供免费使用或试用。下面要介绍的是在 Linux 上使用最广泛的 MySQL 数据库，这也是一种免费的数据库管理系统。

1．MySQL 简介

MySQL 是一个真正的多用户、多线程 SQL 数据库服务器。SQL（结构化查询语言）是世界上最流行的和标准化的数据库语言，它使得存储、更新和存取信息更容易。MySQL 是以一个客户机/服务器结构来实现，它由一个服务器守护程序 mysqld 和很多不同的客户程序和库组成。

MySQL 主要目标是快速、健壮和易用，它能处理与任何可不昂贵硬件平台上提供数据库的厂家在一个数量级上的大型数据库，但速度更快。

MySQL 当前最大的不足是既不支持事务也不保证参考完整性；对事务存取数据库表能明确地为锁定和解锁。

MySQL 和 Apache[8]以及 PHP[9]是 Linux 系统下 Web 服务器、数据库管理系统和动态页面生成的最佳组合。

2．MySQL 的一些重要特征[10]

MySQL 数据库具有以下特征：

- 使用核心线程的完全多线程，这意味着当可能时它能很容易地利用多 CPU。
- 具有多种编程语言的 API，包括 C、C++、Eiffel、Java、Perl、PHP、Python 和 TCL API。
- 可运行在不同的平台上，包括所有主流操作系统和硬件平台。
- 具有多种列类型：1、2、3、4 和 8 字节长度的有符号/无符号整数、FLOAT、DOUBLE、CHAR、VARCHAR、TEXT、BLOB、DATE、TIME、DATETIME、TIMESTAMP、YEAR、SET 和 ENUM 类型。
- 利用一个优化的一遍扫描多重连结（one-sweep multi-join）非常快速地进行连结（join）。
- 在查询的 SELECT 和 WHERE 部分支持全部运算符和函数，例如：
  mysql> SELECT CONCAT(first _ name," ", last _ name) FROM tbl _ name WHERE income/dependents>10000 AND age>30;
- 通过一个高度优化的类库实现 SQL 函数库，并且同样快速。
- 全面支持 SQL 的 GROUP BY 和 ORDER BY 子句，支持聚合函数[COUNT()、COUNT(DISTINCT)、AVG()、STD()、SUM()、MAX()和 MIN()]。
- 支持 ANSI SQL 的 LEFT OUTER JOIN 和 ODBC 语法。
- 可以在同一查询中混用来自不同数据库的表。

- 一个非常灵活且安全的权限和口令系统，并且允许基于主机的认证。口令是安全的，因为当与一个服务器连接时，所有的口令传送被加密。
- 具有给 Windows 操作系统家族使用的 ODBC。具有所有的 ODBC 2.5 函数和其他许多函数。
- 具备索引压缩的快速 B 树磁盘表。
- 每个表允许有 16 个索引。每个索引可以由 1～16 个列或列的一部分组成。最大索引长度是 256 个字节(可以在编译 MySQL 时改变)。一个索引可以使用一个 CHAR 或 VARCHAR 字段的前缀。
- 定长和变长记录。
- 用作临时表的内存散列表。
- 可以进行大数据库处理。
- 所有列都有缺省值。
- 为了可移植性，使用 GNU Automake, Autoconf 和 libtool。
- 用 C 和 C++ 编写，并用大量不同的编译器测试。
- 一个非常快速的基于线程的内存分配系统。
- 没有内存漏洞。用一个商用内存漏洞监测程序测试过(purify)。
- 包括 myisamchk，一个检查、优化和修复数据库表的快速实用程序。
- 全面支持 ISO-8859-1 Latin1 字符集。
- 所有数据以 ISO-8859-1 Latin1 格式保存。所有正常的字符串比较是忽略大小写的。
- 根据 ISO-8859-1 Latin1 字符集进行排序(目前瑞典语的方式)。通过在源代码中增加排序顺序数组可以改变它。MySQL 支持可在编译时指定的很多不同的字符集。
- 表和列的别名符合 SQL92 标准。
- DELETE、INSERT、REPLACE 和 UPDATE 返回有多少行被改变(受影响)。
- 函数名不会与表名或列名冲突。对函数调用的惟一限制是函数名与随后的"("不能有空格。
- 所有 MySQL 程序可以用选项 help 或? 获得联机帮助。
- 服务器能为客户提供多种语言的出错消息。
- 客户端使用 TCP/IP 连接或 Unix 套接字(socket)或 NT 下的命名管道连接 MySQL。
- MySQL 特有的 SHOW 命令可用来检索数据库、表和索引的信息，EXPLAIN 命令可用来确定优化器如何解决一个查询。

第三节　蛋白质组研究中常用的网站及数据库

从 1994 年起每年第一期《核酸研究》(*Nucleic Acids Research*)是分子生物学数

据库专刊，由 Andreas D. Baxevanis 综述当前的在线分子生物学数据库资源。本节介绍蛋白质组研究中常用的网站及站点。

一、蛋白质数据库

1. SWISS-PROT/TrEMBL[11]

网址：http://www.expasy.ch/sprot/ 或 http://www.ebi.ac.uk/swissprot/

内容：注释的(Curated)蛋白数据库

简介：SWISS-PROT 高度注释(比如蛋白功能描述、结构域结构、转录后修饰、变异等)，冗余程度最低，与其他数据库整合程度最高。TrEMBL 是 SWISS-PROT 的补充，含有所有的 EMBL 核苷酸的翻译产物，但未整合进 SWISSPROT。

SWISS-PROT 是一个准确的数据库，它努力地提供高度的注释(比如蛋白功能的描述、它的结构域结构、转录后修饰、变异度等)，最低的冗余和与其他数据库的高度整合。最近数据库的进展包括格式和内容的加强，与其他数据库的交叉引用，新的文档以及对 TrEMBL 的改进，由计算机进行地对 SWISS-PROT 的增补。TrEMBL 是源于 EMBL 核酸序列数据库中的所有编码序列(CDS)的翻译产物(除非已经包括在 SWISS-PROT 中的 CDS)，由 SWISS-PROT 样格式的条目组成。

2. Protein Information Resource (PIR)[12]

网址：http://pir.georgetown.edu/

内容：广泛的、注释的、非冗余的蛋白质序列数据库

简介：蛋白信息资源(Protein Information Resource)，与慕尼黑蛋白序列信息中心(Munich Information Center for Protein Sequences, MIPS)及日本国际蛋白质信息数据库(Japan International Protein Information Database, JIPID)合作，产生了公共领域中最广泛的熟练注释的蛋白序列数据库，PIR 国际蛋白质序列数据库。为了提供及时的、高质量的注释并且提高数据库的可互作性，国际 PIR 在受约词汇(controlled vocabulary)和标准化命名上用基于规则和基于分类驱使的程序，同时也包括了由预测的蛋白质特征决定的实验差别的状态标签。此数据库含有 250 000条以上非冗的蛋白质序列，被划分成家族和超家族，其结构域和基序也被鉴定。其条目与其他序列、分类、基因组、结构和活性数据库广泛地交叉应用。PIR 网站的特征搜索引擎用序列相似性和数据库注释来帮助蛋白的分析和功能鉴定。PIR 过集数据库和搜索工具可以在 PIR 网站(http://pir.georgetown.edu/)上或慕尼黑网站(http://www.mips.biochem.mpg.de)上得到。PIR 国际蛋白质序列数据库和其他文件也可以通过 FTP 得到。

PIR-PSD 是通过 Oracle 对象-关系数据库实现的，而且可得到具有 DTD 的 XML 格式文档。

3. NCBInr

是一个非冗余的蛋白质数据库，由 NCBI 搜集，以供其搜索工具 BLAST 和 Entrez 所用。这些条目(entry)从 GenBank CDS 转录物、PIR、SWISS-PROT、PRF 和 PDB 中搜集得到。其访问地址是：http://www.ncbi.nlm.nih.gov.

4. dbEST

是 GenBank 的分支，含有从大量的生物中来的一次反应测序(single-pass)得到的 cDNA 或叫表达序列标签。这是一个核酸数据库，被 Mascot 从 6 个相位翻译得到。

5. OWL

OWL 混合蛋白质数据库(Composite Protein Sequences Databases)是一个非冗余的蛋白质序列数据库，由 4 个公用的一级资源组成：SWISS-PROT、PIR (1-3)、GenBank (translation)和 NRL-3D。其访问地址是：http://www.biochem.ucl.ac.uk/bsm/dbbrowser/OWL/owlcontents.html.

6. UniGene

美国国家生物技术信息中心（NCBI）提供的公用数据库，该数据库将 GenBank 中属于同一条基因的所有片段拼接成完整的基因进行收录。

二、蛋白质组数据库

1. AAindex[13]：氨基酸索引数据库 (Amino Acid index database)

网址：http://www.genome.ad.jp/dbget/

内容：肽的物理化学属性

简介：AAindex 是一个氨基酸索引(amino acid index)和氨基酸突变矩阵(amino acid mutation matrix)数据库。氨基酸索引是一套代表各种氨基酸的物理化学和生物化学属性的 20 个数值。氨基酸突变矩阵一般是一个代表氨基酸相似性的 20×20 的数值矩阵。Aaindex 由两部分组成：AAindex1 搜集发表的氨基酸索引，AAindex2 搜集发表的氨基酸突变矩阵。AAindex1 或 AAindex2 的条目由以下几部分组成：定义、参考信息、相关系数的列表、实际数据。此数据库可以通过 GenomeNet (http://www.genome.ad.jp/aaindex/)上的 DBGET/LinkDB system 进入，也可以通过匿名 FTP (ftp://ftp.genome.ad.jp/db/genomenet /aaindex/)下载。

2. GELBANK[14]

网址:http://gelbank.anl.gov

内容:全基因组的二维凝胶电泳图谱

简介:GELBANK 是一个公用的数据库,搜集了已知基因组信息的生物的蛋白质组二维凝胶电泳凝胶图(http://gelbank.anl.gov)。它包括 NCBI 已完成的微生物和人类蛋白质组,带有一个搜索引擎,可以通过描述相对分子质量、等电点和蛋白的序列信息快速检索序列。每一个蛋白质组均提供了可浏览的理论上的 2D 胶图谱。GELBANK 允许注册用户上载 2D 胶图谱和它们的注释。现在,此网站提供了 *Methanococcus jannaschii*、*Escherichia coli*、以及 *Deinococcus radiodurans* 的 2D 胶图谱。通过应用 web 界面,提出了一种展示与蛋白质组数据库交互和整合的 2D 胶图谱的方法。

3. Predictome[15]

网址:http://predictome.bu.edu

内容:预测蛋白质间功能关系的数据库

简介:当前基因组序列的泛滥导致了某些分析工具的出现,这些工具致力于搞清这些数据的含义。为了得到蛋白质间的关系,计算和高通量的实验方法出现了。这些方法在提出假设方面非常有效,允许研究者扫描大量的数据以得到有兴趣的、可供进一步详细研究的图谱。虽然在基因功能预测中这些预测的关系的潜在用途已经被证明,但对许多蛋白质组的所有的这种联系的搜集可以使其用途达到最大化。为此 Predictome 出现了,这是一个预测蛋白质间关系的数据库,这些蛋白质间的关系是基于将三种计算机预测方法(即染色体相邻法、系统发育谱法、结构域融合法)应用在 44 个基因组上以及大规模实验方法扫描蛋白质间相互作用而得到的。这些从各种预测方法中得到的数据在一个数据库中的组合使我们可以对它们进行相互间的比较,也可以看到它们与已知通路的关系。作为这些数据的仓库,Predictome 将成为公共资源,为新产生的基因组数据提供蛋白质间的功能关系。

4. Proteome Analysis Database[16]**:蛋白质组分析数据库**

网址:http://www.ebi.ac.uk/proteome/

内容:InterPro 和 cluSTr 在全基因组的蛋白质功能分类中的在线运用

简介:蛋白质组分析数据库的创立提供了对基因组测序完成的生物(包括真细菌、古细菌和真核生物)预测的蛋白质组的广泛的统计和比较分析。此分析是在 InterPro、CluSTr 和 GO 上进行的,在非冗余的 SWISS-PROT 和 TrEMBL 的完全的蛋白质组的条目中执行。InterPro 和 CluSTr 为家族、结构域、位点提供

了一个新的视野，覆盖了每个全蛋白质组的40%到70%（InterPro统计）的蛋白质。Proteome Analysis Database包括一个程序，该程序是专门执行InterPro蛋白质组比较的，比较可用任何一个蛋白质组对数据库中另外一个或多个蛋白质组进行。此数据库的另一特征是提供链接，从HSSP、PDB和SCOP得到每个蛋白质组的结构信息，以及从GO得到功能实体（ontology）。

近期许多新的蛋白质组已经加入，现在此数据库含有53个蛋白质组。人、小鼠的全部编码序列的预测结果在EMBL核酸序列数据库、SWISS-PROT和Ensembl（http://www.ensembl.org）没有提供，但它们共同提供了这些物种蛋白质组草图。这些数据作为蛋白质组分析数据库的一部分提供，包括对这些数据集的基于InterPro的统计和比较分析。其他的进展包括：蛋白长度分布的统计、到SCOP和NEWT的链接、分类数据库的浏览（http://www.ebi.ac.uk/webapps/taxonomy/frameset.html）。

5. REBASE[17]

网址：http://rebase.neb.com/rebase/rebase.html

内容：限制酶和相关的甲基酶

简介：REBASE包含有关限制酶、DNA甲基酶和相关蛋白（如切刻酶、特征性的亚基和控制蛋白）的信息。它包含了已发表的和未发表的参考文献，识别和切割位点，同裂酶，商业效用，甲基化敏感性，晶体结构数据和序列数据，也包括归巢内切核酸酶（homing endonucleases）。最近，关于限制酶的甲基化敏感性信息也加入，并且还含有对测序完毕的真细菌和古细菌的基因组的推测限制酶系统的分析结果。这些数据通过e-mail发布，ftp地址为ftp://ftp.neb.com，网站为http://rebase.neb.com。

6. SWISS-2DPAGE[18]

网址：http://www.expasy.org/ch2d/

内容：注释的二维聚丙烯酰胺凝胶电泳数据库

简介：SWISS-2DPAGE是一个1993年建立的注释的二维聚丙烯酰胺凝胶电泳数据库。现在的版本包括24个参考图，是从人和鼠的生物样本以及最初的啤酒酵母（*Saccharomyces cerevisiae*）、大肠杆菌（*Escherichia coli*）和盘基网柄菌属（*Dictyostelium discoideum*）中得到。此参考图现在有2824个鉴定点，对应于除了各条SWISS-2DPAGE序列虚拟的或任何用户输入的氨基酸序列条目外的数据库中614个独立的蛋白条目。去年SWISS-2DPAGE数据库有如下进展：加入了三个新图，好几个其他的图被更新；加入了与新建的同盟2-DE数据库的交叉索引；新的访问此数据的功能通过ExPASy蛋白质组服务器提供。

SWISS-2DPAGE数据库在1993年建立，由Central Clinical Chemistry Labo-

ratory of the Geneva University Hopital 和 Swiss Institute of Bioinformatics (SIB) 共同维护。现在（2001 年 10 月）SWISS-2DPAGE 数据库的版本包括 31 个参考图 781 个条目，这些参考图的样本源于人类（肝、血浆、HepG2、血红细胞、淋巴瘤、HepG2 分泌蛋白、脑脊液、巨噬样细胞系、红白血病细胞、血小板、肾、早幼粒白血细胞、结肠上皮细胞、结肠腺癌细胞系（DL-1），肝组织来源的可溶性核蛋白和基质），小鼠（肝、腓肠肌、胰岛细胞、白、褐脂肪组织、肝组织来源的可溶性核蛋白和基质），拟南芥菜，大肠杆菌（7 个等电点范围：3.5～10，4～5，4.5～5.5，5～6，5.5～6.7，6～9，6～11），酵母和盘基网柄菌属（*Dictyostelium discoideum*）。在这些图上的 3326 个点已经通过凝胶匹配（47.7%）、免疫印迹（20.5%）、微量测序（15.8%）、氨基酸组成（5.4%）、共迁移（2.6%）和肽质量指纹谱（24.8%）鉴定出来了。有些点是用多种方法鉴定的，因此上述总数超过 100%。通过 ExPASy 服务器，SWISS-2DPAGE 数据库可以通过关键词［蛋白的描述、蛋白名、基因名、物种、作者、全文、蛋白点流水号（serial number）］检索或通过点击图形上的点来查询。也可以计算 SWISS-2DPAGE 数据库中每一个蛋白序列或用户输入的氨基酸序列在 SWISS-2DPAGE 参考图中的估算的定位。

7. YPL. db[19]

网址：http://ypl.tugraz.at

内容：酵母蛋白定位数据库

简介：Yeast Protein Localization database（YPL.db）含有由显微分析（microscopic analyses）得到的酵母蛋白定位图谱信息。得到这些定位信息的实验数据和参数以及从激光共聚焦（confocal）或影像显微镜（video microscopy）得到的图像均存储于一个相关的数据库，作为所有实验的存档和证明。此数据库可以通过基因名、蛋白定位、生长环境和大量的附加参数来检索。所有实验参数可以从预先给定的列表中选择以确保对不同的研究者的数据库的完整性和一致性。此数据库提供了一个结构参考资源，可以对未知和不明确的定位图谱作更好的判定。与 MIPS、YPD 和 SGD 数据库的链接可快速访问在此定位数据库中不包含的进一步的信息。

三、蛋白质序列基序（motif）

1. Blocks[20]

网址：http://blocks.fhcrc.org

内容：蛋白家族保守区的多重比对

简介：Blocks 是一个没有空位的多重比对，对应于蛋白质最保守的区域。

Blocks 数据库是由 Blocks 组成，这些 Blocks 是用自动的 PROTOMAT 系统从相关蛋白家族中提取构建。该数据库现在含有 11 853 个 Blocks，代表了 2 608 个在 InterPro 和 Prints 中搜集的蛋白质家族。一个 Blocks 的多重比对由没有空位的保守区组成，这些保守区被不能比对上的可变长度的区域分割开。PROTOMAT 系统用一个强有力的 motif-finder 对一组相关蛋白序列进行搜索。结果得到的候选基序被组装成沿着序列排列的一个最佳的集合，其他序列如果已知与一个家族有关并且能够与其所有的 Blocks 结果比对上则也可加入。为了降低 Blocks 数据库的冗余度和大小，当文档质量被认可而又缺少家族之间的关系时，PROTOMAT 可以按分层方式将其应用到 InterPro 的序列家族中。由于 Prints 的数据库格式和 Blocks 的一样，Prints 的 Blocks 可以不用运行 PROTOMAT 而直接加入。LAMA 算法被用来将所有加入到 Blocks 数据库的 Blocks 与其余的相比较，以减少冗余的条目。Blocks 数据库被用来注释未知功能的蛋白质。蛋白或 DNA 序列可以用 BLIMPS 搜索工具或 IMPALA 以及 NCBI 的 RPS-BLAST 工具与 Blocks 数据库比较，所有的这些工具提供了对命中的条目的统计评估。当一个对 Blocks 数据库的搜索命中了一个蛋白质家族，用户就被链接到 InterPro 页面上。Blocks 的 www 服务器也提供了几种可以增强所获得的信息的工具。一个家族的 Blocks 的图形化显示(visual displays)可由序列 logos 沿着序列的 Blocks 图以及它们中已知结构序列的 Blocks 来提供。由代表每一个家族的 Blocks 构建的系统发生树能用 ProWeb 树形显示来探索。对序列数据库的 Block 比对的反向搜索可以通过链接到 BLAST、MAST 和 LAMA 搜索网页完成。还提供有到同样的匹配 Blocks 的 CYRCA 集的链接，这个匹配的 Blocks 常常是在不同的背景中出现的具有相似的结构和功能的蛋白区段。CYRCA 集是使用 LAMA 算法通过 Blocks 的成对比较找到的。Blocks 能链接到 CODEHOP PCR 引物设计工具，此引物设计工具用多重比对来设计杂交共有的简并引物(hybrid consensus-degenerate primers)。最后，Blocks 能链接到 SIFT，SIFT 是一个从多重比对信息中预测氨基酸替换效果的程序。当用户在用 Block Maker 和 Multiple Alignment Processor 方法从自己的序列中得到 Blocks 或从自己的多重比对序列中切出 Blocks 时，这些工具也很有用。

InterPro 和 Prints 是现在 Blocks 数据库的蛋白质家族的仅有资源，Blocks 是在 2001 年 8 月从 InterPro 3.1 和 Prints 31.0 中得到更新。由一个家族的蛋白序列的保守区得到的系统发生树现在能用 ProWeb 树形显示来探索。它允许交互式的缩放，而且可以从相关的亚树中得到新的 Blocks。当原来的 Blocks 家族含有明显的亚家族时，这是很有必要的。Blocks 现在已被链接到有一样比对的 Blocks 的 CYRCA 集中。CYRCA 集是通过用 LAMA 算法的成对的 Blocks 的对排而得到。可对排的 Blocks 常常是在不同的背景中有相似的结构和功能的蛋白质区域。CYRCA 有一个能将属于此集的序列的结构重叠的工具。Blocks 可以用 SIFT 分析，SIFT 是一个用多重比对预测是否一个氨基酸的替换会影响蛋白质功能的工

具。SIFT 利用了一个事实，那就是保守位点的替换，如在 Blocks 区，更有可能影响功能。

2. CDD[21]

网址：http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml

内容：保守蛋白结构域的比对模型

简介：保守结构域数据库(Conserved Domain Database, CDD)是一个代表在分子进化中蛋白保守结构域的多序列比对谱。已经通过公开搜集的 Pfam 和 Smart 比对数据以及 NCBI 成员提供的数据组装完成。当前 CDD 版本(v1.53)含有 3551 个这种模型。CDD 的比对被链接到 Entrez 的蛋白序列和结构数据中。分子结构显示软件 Cn3D 是一个可以交互式地显示比对和三维结构，并由进化保守特征注释三维残基坐标的工具。CDD 可以通过万维网进入，其 URL 为 http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml。蛋白质查询序列可用一个叫 CD 搜索的服务(http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)对位点特异打分矩阵(PSSMs)数据库进行比较。位点特异打分矩阵是从 CDD 的比对中得到的。CD 搜索运行反向位点特异的 BLAST(reverse position-specific BLAST, RPS-BLAST)，是一个使用广泛的 PSI-BLAST 算法的变体。CD 搜索是提交到 NCBI 的蛋白质-蛋白质间 BLAST 查询服务(http://www.ncbi.nlm.nih.gov/BLAST)的默认运行程序。

3. CluSTr[22]

网址：http://www.ebi.ac.uk/clustr/

内容：对 SWISS-PROT + TrEMBL 蛋白质的自动分类

简介：CluSTr (Clusters of SWISS-PROT and TrEMBL proteins)数据库可自动对 SWISS-PROT 和 TrEMBL 蛋白质进行分类，分成几个相关蛋白群。此分类是基于对蛋白质序列间所有的两两比对的分析。此分析分别对不同蛋白质相似性水平进行，产生一个成簇的类别。这个数据库提供了与 InterPro 的链接，InterPro 整合了从 PROSITE、PRINTS, Pfam、ProDom 和 SMART 来的蛋白质家族、结构域和功能位点信息。与 InterPro 的图形化界面的链接使用户一眼就能看出是否此簇中的蛋白共有特定的功能位点。CluSTr 也提供了与 HSSP 和 PDB 的交叉索引。

过去一段时间的工作集中在为全蛋白质组创立簇。42 个完成的原核蛋白质组的簇已建立。除了已有的真核蛋白质组(线虫、啤酒酵母、果蝇)外，现又加入了拟南芥的数据。由于人和鼠的预测的完全编码序列在 EMBL 核酸数据库中未提供，给出了从 SPTr 和 Ensembl (http://www.ensembl.org)全蛋白质草图来的 SWISS-PROT 和 TrEMBL 的蛋白簇。

4. InterPro[23]

网址:http://www.ebi.ac.uk/interpro/

内容:蛋白质家族、结构域、位点的整合资源

简介:带有蛋白质家族、结构域或功能位点的鉴别信号(signatures diagnostic)的数据库对于新鉴定的缺乏生化特征的序列的计算机分类是一个重要的工具。InterPro是一个蛋白质家族、结构域和功能位点的整合资源,它合并了PROSITE、PRINTS、Pfam、ProDom以及SMART数据库的成果。每个InterPro条目包括功能描述、注释、引用文献、与相关的子数据库的链接。一个新的特征是包括InterPro条目到相关的Gene Ontology (GO)条目的映射,它是一个从蛋白质到GO terms的自动映射的一个有用的工具。InterPro的3.2版(2001年7月)含有3 939个条目,代表了2850个家族、1009个结构域、65个重复序列和15个转录后修饰(PTMs)位点。每个InterPro条目列出了与SWISS-PROT和TrEMBL的所有匹配(在SWISS-PROT和TrEMBL的约580 000条蛋白中有2 000 000个命中)。下一个子数据库TIGRFAMs的整合现在正在进行中。该数据库可通过文本和基于序列的检索来访问。对该数据库有何疑问可e-mail至:Interhelp@ebi.ac.uk.

现在,从子数据库中得到的新的鉴别信号已被整合进InterPro。SMART (a Simple Modular Architecture Research Tool)数据库已经被整合进InterProt,SMART数据库允许遗传上可移动的结构域的鉴定和注释以及结构域构件的分析。它包括500多个在信号、细胞外和染色质相关蛋白中的结构域家族。现在正在进行的子数据库的整合,TIGRFAMs,提供了隐马尔可夫模型,以确定同一功能的蛋白质家族。其他新特征包括一个附加的图形化用户界面,以在一个浓缩的图形显示(condensed graphical view)中显示蛋白质匹配,以及与Gene Ontology (GO)条目的映射。可能的话,InterPro的条目已经手工GO条目化,以使这些条目可应用到所有的与这条目匹配的蛋白中。

5. Pfam[24]

网址:http://www.sanger.ac.uk/Software/Pfam/

内容:多序列比对和共同的蛋白质结构域的隐马尔可夫模型

简介:Pfam是一个大的蛋白多序列比对集和隐马尔可夫模型的profile。Pfam以万维网形式由英国的http://www.sanger.ac.uk/Software/Pfam/、瑞典的http://www.cgr.ki.se/Pfam/和美国的http://pfam.wustl.edu/提供。Pfam的最新版本(4.3)含有1815个家族。Pfam的家族与SWISS-PROT 37和TrEMBL 9上的蛋白有63%是匹配的。现在Pfam对于全基因组有超过半数的蛋白能对应上。基因组DNA能直接用Wise2软件包对Pfam库进行搜索。

6. PROSITE[25]

网址：http://www.expasy.org/prosite

内容：有生物学意义的蛋白 patterns 和 profiles

简介：PROSITE(http://www.expasy.org/prosite/)是由有生物学意义的 pattern 和 profile 组成，用适当的计算工具，它可以有助于判断一个新的序列属于哪个家族(如果存在的话)，或含有哪些已知的结构域。

现在 profile 的搜集扩展到组成上的偏性区(如富含 Pro)和所谓的环形 profiles，环形 profiles 可在一个由几个串联重复单元组成的重复区产生一个匹配(如 COLLAGEN-REP)。一个新的 profile 搜索服务器(http://hits.isb-sib.ch/cgi-bin/hits-motifscan)的建立提供了一个对其收集(patterns、profiles、pfam-hmms)的每周自动更新及与 Hits 数据库(http://hits.isb-sib.ch)的链接。

7. 其他蛋白质序列基序数据库(表 9-1)

表 9-1 其他蛋白质序列基序数据库

数据库	网址	说明
EMOTIF	http://motif.stanford.edu/emotif	蛋白序列基序的判断和搜索
iPROCLASS	http://pir.georgetown.edu/iproclass/	用结构和功能信息注释的蛋白质分类数据库
O-GLYCBASE	http://www.cbs.dtu.dk/databases/ OGLYCBASE/	糖蛋白和 O-糖基化位点
PIR-ALN	http://pir.georgetown.edu/pirwww/dbinfo/piraln.html	蛋白序列比对
PRINTS	http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/	等级基因家族指纹
ProClass	http://pir.georgetown.edu/gfserver/proclass.html	由PIR 超家族和 PROSITE pattern 定义的蛋白质家族
ProDom	http://www.toulouse.inra.fr/prodom.html	蛋白质结构域家族

续表

数据库	网址	说明
ProtoMap	http://protomap.cornell.edu	自动的 SWISS-PROT 蛋白的自动的等级分类
SBASE	http://www.icgeb.trieste.it/sbase	注释的蛋白质结构域序列
SMART	http://smart.embl-heidelberg.de	简单模块构件搜索工具
SUPFAM	http://pauling.mbu.iisc.ernet.in/~supfam	结构相关的序列家族
SYSTERS, GeneNest, SpliceNest	http://emb.molgen.mpg.de	蛋白质家族、EST 簇和基因组位点的整合数据库

四、蛋白质三维结构和相关数据库

1. PDB[26]

网址：http://www.rcsb.org/pdb/

内容：用 X 射线晶体学和核磁共振法(NMR)得到的结构数据

简介：PDB(Protein Data Bank)是专一的全世界的生物大分子结构数据库。现在产生了 PDB 存档的 mmCIF 数据文件集(ftp://beta.rcsb.org/pub/pdb/uniformity/data/mmCIF/)。同时一个可以将 mmCIF 数据文件转化成 PDB 格式的应用程序(叫 CIFTr)也已公布，以支持已有的软件。

2. MMDB[27]

网址：http://www.sander.ebi.ac.uk/hssp/

内容：所有实验测定的三维结构，链接到 NCBI 的 Entrez

简介：现在，许多蛋白质家族的三维结构已经知道，因此，很有可能在搜索序列数据库时会碰到已知结构的同源物。Entrez 的三维结构数据库的目标就是使这些信息以及其提供的功能注释能够很方便地提供给分子生物学家。为了达到此目标，Entrez 的搜索引擎提供了三个强有力的特征：①序列和结构的邻接(neighbors)；比如，人们可能会选与一个感兴趣的序列相似的所有的序列，以及到所有已知的三维结构的链接。②数据库间的链接；比如，人们可能会通过词汇的匹配搜索 MEDLINE，同时链接到这些文章中所报道的 3D 结构。③序列和结构的可视化；在确定一个已知结构的同源物后，人们可以通过观察分子图像和比对情况来推导近似的 3D 结构。MMDB(Molecular Modeling Database)可以从以下网址得到：http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Structure。

3. 其他有关数据库

BioMagResBank——从核磁共振光谱学(NMR spectroscopy)来的蛋白质、肽及核酸的数据的数据库

SWISS-MODEL Repository——自动产生的蛋白模型数据库

ModBase——比较的蛋白结构模型数据库

CATH——蛋白的 CL BSM 结构分类

SCOP——蛋白的结构分类[在美国、以色列、新加坡、澳大利亚有镜像]

Mol _ R _ Us——Molecules R Us, PDB 的浏览器

BMM Domain Server——生物分子模建实验室(Biomolecular Modelling Laboratory, ICRF) 蛋白结构域服务器

PRESAGE——结构基因组学联合资源

ReLiBase——受体/配基复合体数据库[美国有镜像]

TOPS——蛋白质拓扑图

CCD——剑桥晶体学数据中心(剑桥结构数据库, CSD)

五、在线蛋白工具

表 9-2 介绍了常用在线蛋白工具[28]。

表 9-2 常用在线蛋白工具

软件名称	网址	简介
BCM Search Launcher	http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html	蛋白序列二级结构预测综合站点,从此出发,输入蛋白序列,可以根据需要,使用各种在线预测工具,包括 Coils、nnPredict、PSSP/SSP、PSSP/NNSSP、SAPS、TMpred、SOUSI、Paircoil、Protein Hydrophilicity/Hydrophobicity Search、SOPM,使用十分方便
DAS	http://www.biokemi.su.se/~server/DAS/	蛋白跨膜预测服务器,输入蛋白序列,预测跨膜区域
TopPred 2	http://www.biokemi.su.se/~server/toppred2/	斯德哥尔摩大学理论化学蛋白预测服务器提供的膜蛋白拓扑学预测 Topology prediction of membrane proteins 在线工具
SOSUI	http://azusa.proteome.bio.tuat.ac.jp/sosui/	东京农业科技大学(Tokyo University of Agriculture and Technology)提供的膜蛋白分类和二级结构预测在线工具

续表

软件名称	网址	简介
PSIpred-MEM-SAT2	http://globin.bio.warwick.ac.uk/psipred/	本蛋白结构预测服务器允许您提交一个蛋白序列，进行二级结构预测与跨膜拓扑结构预测，并将结果用 EMAIL 提供给您
HMMTOP	http://www.enzim.hu/hmmtop/	预测蛋白序列的跨膜螺旋与拓扑结构服务器
TMpred	http://www.ch.embnet.org/software/TMPRED_form.html	预测蛋白序列跨膜区
TMHMM	http://www.cbs.dtu.dk/services/TMHMM-1.0/	预测蛋白的跨膜螺旋
The PredictProtein server	http://www.embl-heidelberg.de/predictprotein/	提供蛋白数据库查询，预测蛋白各种结构的服务
SMART	http://smart.embl-heidelberg.de/	提供蛋白序列，在结构域数据库中查询，显示出其结构域及跨膜区等
SPLIT	http://pref.etfos.hr/split/	膜蛋白二级结构预测服务器
PRED-TMR	http://o2.db.uoa.gr/PRED-TMR/	提供基于 SwissProt 数据库统计分析的预测蛋白跨膜片段的服务
CoPreThi	http://o2.db.uoa.gr/CoPreThi/Main.html	基于 INTERNET 的 JAVA 程序，预测蛋白的跨膜区
TMAP	http://www.embl-heidelberg.de/tmap/tmap_info.html	提供预测蛋白跨膜片段的服务
multalin	http://www.toulouse.inra.fr/multalin.html	蛋白序列比对服务器，比较几条蛋白序列的结构
Protein Sequence Analysis	http://bioweb.pasteur.fr/seqanal/protein/intro-uk.html	巴斯德研究所提供的常用蛋白序列分析在线工具
PSA	http://bmerc-www.bu.edu/psa/	Protein Sequence Analysis 服务器为美国波士顿大学生物分子工程研究中心(the BioMolecular Engineering Research Center)开发，提交氨基酸序列，预测二级结构及折叠区域
PRS	http://www.embl-heidelberg.de/prs.html	EMBL 提供的以未知蛋白序列的氨基酸组成而非氨基酸序列顺序进行蛋白家族及各种特性预测的服务器，并将结果通过 E-MAIL 发给您
AAA	http://www.embl-heidelberg.de/aaa.html	EMBL 氨基酸分析服务器，与上一个服务类似，以氨基酸残基组成为蛋白分析基础数据，结合蛋白数据库进行分析

参考文献

1 生命科学发展战略. http://www.kepu.gov.cn/kjqy/file/0204.htm

2 陈润生. 生物信息学. http://www.cpus.gov.cn/kjqy/file/0296.htm

3 后基因组时代序幕拉开. http://www.kepu.gov.cn/kjqy/file/0126.htm

4 张春霆. 生物信息学的现状与展望. http://tubic.tju.edu.cn/current and prospect.htm

5 中华人民共和国科学技术部. 生物和现代农业技术领域生物信息技术主题课题申请指南. http://www.863.org.cn/863_95/applygide/app006_03.html

6 钟扬, 张亮, 赵琼, 简明生物信息学. 北京: 高等教育出版社, 2001, 179～199

7 http://www.bioinfochina.com/resource/ view_chapter.php? id=400

8 http://www.apache.org

9 http://www.php.net

10 http://www.mysql.com

11 Bairoch A and Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Research, 2000, 28: 45～48

12 Cathy H W, Huang HZ, Arminski L, Castro-Alvear J, Chen Y X, Hu Z Z, Robert S L, Kali C L, Hans-Werner M, Bruce C O, Baris E S, Tsugita A, Vinayaka R C, Yeh L S, Zhang J and Barker W C. The Protein Information Resource: an integrated public resource of functional annotation of proteins. Nucleic Acids Research, 2002, 30: 35～37

13 Kawashima S, Kanehisa M. AAindex: Amino Acid index database. Nucleic Acids Res, 2000, 28(1): 374

14 Babnigg G, Giometti C S. http://www3.oup.co.uk/nar/database/summary/222

15 Mellor J C, Yanai I, Clodfelter K H, Mintseris J, DeLisi C. Predictome: a database of putative functional links between proteins. Nucleic Acids Res, 2002, 30(1): 306～309

16 Apweiler R, Biswas M, Fleischmann W, Kanapin A, Karavidopoulou Y, Kersey P, Kriventseva EV, Mittard V, Mulder N, Phan I, Zdobnov E. Proteome Analysis Database: online application of InterPro and CluSTr for the functional classification of proteins in whole genomes. Nucleic Nucleic Acids Res, 2001, 29(1): 44～48

17 Roberts R J, Macelis D. REBASE restriction enzymes and methylases. Nucleic Acids Res, 2001, 29(1): 268～269

18 Hoogland C, Sanchez J C, Tonella L, Binz P A, Bairoch A, Hochstrasser D F, Appel R D. The 1999 SWISS-2DPAGE database update. Nucleic Acids Res, 2000, 28(1): 286～288

19 Habeler G, Natter K, Thallinger G G, Crawford M E, Kohlwein S D, Trajanoski Z. YPL.db: the Yeast Protein Localization database. Nucleic Acids Res, 2002, 30(1): 80～83

20 Henikoff J G, Pietrokovski S, Greene E A, Ng P C, Henikoff S. Increased coverage of protein families with the Blocks Database servers. Nucleic Acids Research, 2000, 28: 228～230

21 Marchler-Bauer A, Panchenko A R, Shoemaker B A, Thiessen P A, Geer L Y, Bryant S H. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Research, 2002, 30: 281～283

22 Kriventseva E V, Servant F, Fleischmann W, Zdobnov E V, Apweiler R. CluSTr: a database of clusters of SWISS-PROT + TrEMBL proteins. Nucleic Acids Research, 2001, 29: 33～36

23 Apweiler R, Attwood T K, Bairoch A, Bateman A, Biswas M, Bork P, Bucher P, Copley R, Croning M D

R, Durbin R, Falquet L, Fleischmann W, Gouzy J, Griffiths-Jones S, Hermjakob H, Hulo N, Jonassen I, Kahn D, Kanapin A, Karavidopoulou Y, Lopez R, Mulder N J, Oinn T M, Pagni M, Ponting C P, Servant F, Sigrist C J A, Zdobnov E M. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Research, 2001, 29: 37～40

24 Bateman A, Birney E, Cerruti C, Durbin R, Etwiller L, Eddy S R, Griffiths-Jones S, Kevin L H, Marshall M and Sonnhammer E L L. The Pfam Protein Families Database. Nucleic Acids Research, 2002, 30: 276～280

25 Falquet L, Pagni M, Bucher P, Bairoch A, Hulo N, Sigrist C J A, Hofmann K. The PROSITE database, its status in 2002. Nucleic Acids Research, 2002, 30: 235～238

26 Westbrook J, Feng Z K, Jain S, Bhat1 T N, Thanki1 N, Ravichandran1 V, Gilliland G L, Bluhm W, Weissig H, Greer D S, Bourne P E and Berman H M. The Protein Data Bank: unifying the archive. Nucleic Acids Research, 2002, 30: 245～248

27 Wang Y, Anderson J B, Chen J, Geer L Y, He S Q, Hurwitz D I, Liebert C A, Madej T, Marchler G H, Marchler-Bauer A, Panchenko A R, Shoemaker B A, Song J S, Thiessen P A, Yamashita R A and Bryant S H. MMDB: Entrez's 3D structure database. Nucleic Acids Research, 2002, 30: 249～252

28 http://biosoft.biosino.org/prool.html

（朱云平）

第十章　细胞器与蛋白质复合体的组成分析

蛋白质组的研究包括应用先进的分离技术、质谱及生物信息学工具去鉴定复杂生物体内的蛋白质。二维凝胶电泳（2-DE）用来分离复杂样品中的蛋白质，为从分子水平研究生物的复杂性提供了有利的工具；蛋白质的肽质量指纹谱（PMF）和蛋白质数据库的搜索常被用来高通量地鉴定从 2-DE 上分离出的蛋白质。随着分析技术的不断提高，过去的 10 年内已经建立了许多不同组织、体液、或细胞系的蛋白质 2-DE 参考图谱。

但是，蛋白质组的研究仍有其缺陷性。其中最为重要的是 2-DE 的分辨率问题，目前 2-DE 的分辨率最多可以达到 3000～4000 个蛋白质点，因此低丰度蛋白质点的分离与鉴定受到限制。在基因组的研究中，低丰度的转录本可以通过抑制消减杂交选择，并用 PCR 进行扩增，而目前在蛋白质组的研究技术中，并没有相关技术，因此许多重要的调节分子，如细胞因子与它们的受体、信号转导蛋白、细胞调节因子及转录因子，通常很难被检测到。另一个困难在于正确地检测疏水性、碱性、相对分子质量极小和极大的蛋白质。此外，蛋白点的定量问题也是一大难题。因此，蛋白质组的研究目前最好以亚细胞的蛋白质为重点（如质膜、细胞器、胞质和核），而不是去描述整个细胞及组织的蛋白质组[1]。

细胞被分成不同的区间，而负责不同细胞机制的功能单位又包含许多大分子结构及多蛋白复合物。从生物学的角度看，真核细胞的区域化作用为细胞中的生化反应提供了可能的调节机制，如 N 端的糖基化只发生在质膜上，因此，只有进入质膜的蛋白质才能发生 N 端的糖基化。为了鉴定整个细胞的蛋白质组，我们可以充分利用细胞的区域化作用和其功能单位，对细胞器及蛋白质复合体的组成进行研究。

第一节　细胞器的分离

一个动物细胞包括很多细胞器，如线粒体、溶酶体、过氧化物酶体、质膜、高尔基体、胞浆及细胞核等，由膜将这些细胞器分隔。不同的细胞系中每个细胞器所占的比例均不相同。

一、线粒体

在成熟的成纤维细胞中线粒体（mitochondria）是含量最丰富的细胞器，每个细胞中约有1000个线粒体，占整个细胞体积的25%。所有线粒体都由一个双层膜系统组成，外膜与内膜分隔开并包围着内膜。线粒体是一种具有能量转换、物质运输、信息传递和合成腺苷三磷酸（ATP）等功能的主要细胞器。

制备线粒体的关键是要保持其完整性及纯度。尽管没有一种方法完全适用于从各种组织、细胞中分离线粒体，但却有一个基本原则，即先制备组织或细胞匀浆，然后进行差速离心，低速去除细胞核及细胞碎片；高速离心分离线粒体，如果有必要，还可以进行密度梯度离心。组织细胞的特异性及所要线粒体的实验用途决定了方法的细节。这些细节包括：匀浆器的选择、缓冲液的成分、许可混杂其他细胞器的程度等。要选择对某种组织来说最合适的缓冲液并确定最好的匀浆方法，在Richwood（1992）的文章中对一些手段以及它们所适用的组织和细胞作了很好的阐述[4]。

大鼠肝脏比较易于得到和匀浆，而且其细胞中含大量线粒体，因此，大鼠肝脏成为分离线粒体最常用的组织。本节以大鼠肝脏为例介绍线粒体的分离方法。

1. 操作程序

1）将饥饿一天的300g左右雄性大鼠处死，迅速取出肝脏，注意不要弄破胆囊。肝脏放进一置于冰上的平皿中，剪去结缔组织，称其质量后放回平皿并浸于缓冲液A中（250mmol/L甘露糖醇、0.5mmol/L EGTA、5 mmol/L Hepes、0.1%（W/V）BSA，pH7.4），用锋利的剪刀将其剪碎，并用缓冲液A冲洗去除大部分的血。转移至匀浆器中，加入足够的缓冲液A进行匀浆。

2）将匀浆转至离心管中离心，600g，5min，2次。取上清。去除细胞核及碎片。

3）上清10 300g离心，10min。沉淀部分为线粒体粗制品，漂洗线粒体两遍。

4）将最终的沉淀用一定体积的适于后续工作的缓冲液悬浮。

5）线粒体的进一步纯化。

线粒体组分可通过密度梯度离心法进一步纯化。通常用Ficoll、Percoll和蔗糖来形成密度梯度。Ficoll和Percoll有商品化的产品。Ficoll是最早发展起来的专门用于细胞分级分离的介质[2,3]。它常用于非连续梯度，普遍用于从脑部分离线粒体和神经突触体。Percoll可使用预先形成的不连续或连续梯度，或在离心过程中使其自动形成梯度，线粒体可在这个梯度中形成等密度带，其他细胞器因密

度较低而位于离心管上部。它广泛用于各种细胞器的分离。其整个梯度都是等渗的，而且能很快地自动形成梯度。然而作为一种胶体悬液，在使用中存在着较难清除这一缺点。到目前为止，非连续蔗糖梯度是最常用的介质，尽管线粒体会被离心时产生的高渗透压破坏，一些代谢功能也会改变或丧失。因为蔗糖比较容易得到，且较为便宜，得到的线粒体可在许多实验中使用[5]。

a. 小心地在 15ml 1.5mol/L 的蔗糖溶液上加一层 15ml 1.0mol/L 的蔗糖溶液，形成 30ml 的梯度。1.0mol/L 的蔗糖和 1.5mol/L 的蔗糖溶液中含：10mmol/L Tris-HCl（pH7.5）；1mmol/L EDTA，pH7.5。

b. 小心加入线粒体悬液，60 000g（22 000 r/min）离心 20min。线粒体会在 1.0mol/L 和 1.5mol/L 界面处形成一薄层。

c. 移去样品层及其与 1.0mol/L 层的交界。

d. 用一巴斯德吸管轻柔地在 1.5mol/L 层顶部吸出线粒体。

e. 稀释蔗糖溶液，17 000 离心 15min，沉淀线粒体。

f. 用 30ml 缓冲液 A 洗一次，重悬浮于合适体积的适于后续工作的缓冲液中。

2. 注意事项

(1) 全部操作均在 0--4℃下进行。

(2) 多次少量制备的总得率要大于一次大量制备的得率，即制备规模的扩大并不意味着得率的提高。

(3) 不要超过梯度容量上样。总的来说，一个梯度可加载来自一只大鼠肝脏或 2--4 组织培养细胞的线粒体

来源是大鼠的肝脏[6]，下面将介绍从大鼠的肝脏中分离高尔基体的过程。

简要地说，方法如下：首先用低剪切力的匀浆处理使细胞破碎，然后通过差速离心浓缩高尔基体。最终的纯化通过蔗糖梯度离心完成，蔗糖梯度离心通常在50 000g 至 100 000g 之间进行[5]。

表 10-1　几种线粒体标志酶及其定位

标志酶	定　位
线粒体的标志酶	
单胺氧化酶	外膜
腺苷酸激酶	中间膜
琥珀酸-细胞色素 C 还原酶	内膜
细胞色素 C 氧化酶	内膜
延胡索酸酶	基质膜
非线粒体标志酶	
葡萄糖-6-磷酸酶	微粒体
酸性磷酸酶	溶酶体
5′-核苷酸酶	质膜
过氧化氢酶	过氧化物酶体

1．操作程序

1）取雄性大鼠（约 250g），用断头放血的方法处死。取肝（10～13g），用刀片切碎。

2）用匀浆器将每份肝脏在 2 倍体积的匀浆介质（50 mmol/L Tris-马来酸，pH6.4；0.5 mol/L 蔗糖；1%葡聚糖）中处理 40s。

3）匀浆液于 5000g 离心 15min，去除大部分上清。用吸管将沉淀中黄褐色的部分重悬于少量的上清中。

4）用吸管反复吹吸使之混匀，将悬液铺在 1.2 mol/L 蔗糖上。

5）样品于 100 000g 离心 30min，吸出上清后，用吸管挑起高尔基体。

6）将这部分高尔基体悬于以下中的一种溶液中：

a. 来自蔗糖梯度的清亮上清，适于最佳形态的保持。

b. 蒸馏水用于达到最高的分布纯度。

c. 匀浆介质或酶分析混合液用于保持酶的活性。

7）以 5000g 离心 15～20min 收集高尔基体。

2. 注意事项

(1) 为了避免高尔基体与溶酶体的长时间接触，组织匀浆随后的步骤应尽可能地快，应尽量缩短高尔基体在匀浆液中非沉淀状态的时间。对于一些肝脏的匀浆液，如果匀浆与离心之间的间隔长于 5min，就会导致 50% 高尔基体的丧失。

(2) 已有许多梯度方法用于高尔基体的纯化。我们选择不连续蔗糖密度梯度方法是因为它的简便性，但连续梯度并不比不连续梯度有明显的优势，在 1.0～1.25mol/L 蔗糖浓度范围内的分步纯度均没有明显的差别。

3. 高尔基体纯度鉴定

用与来源组织相对应的糖基转移酶对高尔基体进行鉴定是最可靠的方法。半乳糖苷转移酶对应于动物，葡聚糖合成酶 1 对应于植物。用于哺乳动物细胞及组织的标志酶在表 10-2 中列出。在确定高尔基体在丰富提取物中的存在及估测它们的丰度时，用电子显微镜鉴定固定及包埋的组织仍然是比较确定和不可替代的方法[6]。用负染来鉴定高尔基体的方法要求较高，定量则更加复杂化。

表 10-2 有助于检测从哺乳动物细胞和组织中所分离高尔基体的标志酶

标志酶	定位
高尔基体的标志酶	
半乳糖苷转移酶	高尔基体
非高尔基体标志酶	
5′-核苷酸酶	质膜污染
K^+-刺激、oubain-抑制的对-硝基苯-磷酸酶	质膜污染
葡萄糖-6-磷酸酶	内质网污染
NADH-细胞色素 C 氧化还原酶	内质网污染
琥珀酸-INT-还原酶	线粒体污染

三、过氧化物酶体

过氧化物酶体（peroxisomes 又称微体）是存在于大多数真核细胞中的亚细胞器，它们通常含有致密的基质，有时还含有一个核心，周围包被着单层膜，大

是一些使 O_2 形成 H_2O_2 而催化代谢物氧化的氧化酶，H_2O_2 则由催化酶分解。过氧化物酶体也参与细胞内脂肪和氨基酸代谢。

有许多方法用于从不同的生物组织中分离过氧化物酶体[7]。这里简要介绍大鼠肝脏中过氧化物酶体的分离[5]。

1. 操作程序

1）乙醚或其他麻醉试剂使大鼠麻醉，采用断头法处死。
2）迅速取出肝脏，剪碎并转移至匀浆器中，加三倍体积（ml/g 组织）的冷匀浆缓冲液匀浆。
3）将匀浆液于 1000g，离心 10 min（4℃）。
4）轻轻倒出上清并保存。再用 3 倍体积的匀浆缓冲液将沉淀进行匀浆。
5）将匀浆液于 600g 离心 10min。
6）将上清与第一次的上清合并；重复 4、5 两步。
7）将两次的上清合并，加入 2 ml PMSF。沉淀（粗的细胞核组分）弃去。
8）将合并后的上清于 3300g 离心 10min，将最后的 1～2 ml 摇晃以尽可能地沉淀上层松软的粉红色部分，同时又不影响下层比较结实的沉淀。弃去上清，用相当于原来体积三分之一的匀浆缓冲液匀浆。
9）将匀浆液于 25 000g 再离心 10min，弃去上清。
10）轻轻地将沉淀重悬于与起始肝脏重量相等体积（ml/g）的匀浆缓冲液中。
11）吸 10ml 30%的 Accudenz 溶液至一只 40ml 的超速离心管中，然后轻轻地将 2ml 重悬的 L 组分铺在 Accudenz 溶液上。盖上管子，45 000g 离心 15min。此外，也可以在冷冻低速离心机中以 25 000g 离心 20min。30%(w/v)的 Accudenz(原来被称作 Nycodeaz，即，5-(*N*-2,3-dihydmxypropyl-acetamino)-2,4,6-triiodo-*N*,*N*′-bis(2,3-dihydmxypmpyl)isopthalamide，Accurate Chemical)溶液，其中含有 10mmol/L 的 TES 及 1 mmol/L EDTA，pH 7.5。）
12）吸出界面物质，接着吸出上清。将沉淀重悬于相当于 1/4 肝脏重量的 Accudenz 介质中（或者悬于匀浆缓冲液中）。这就是最终的过氧化物酶体制备物。Accudenz 可以通过在 4℃对 0.25mol/L 蔗糖、10mmol/L Tris-HCl（pH 7.5）1mmol/L EDTA 透析来除去。

2. 注意事项

1）细胞匀浆液和亚细胞组分应始终保持于 0～4℃，匀浆时匀浆器应浸在冰水中以使管子保持低温，以保持细胞器的完整并尽量减少内源性蛋白酶的降解。

2）应在分离过程中经常加入丝氨酸蛋白酶的抑制剂 PMSF，因为它在水相介质中很容易水解。

3）还应该注意的是，肝脏和肾脏的过氧化物酶体非常脆弱，因此在分离过程中应避免剧烈的匀浆或强烈的渗透压休克，以防止这些细胞器的裂解。最终的过氧化物酶体制备物保存在含有 Accudenz 的高渗溶液中。

4）Accudenz 在 UV 区有强烈的吸收（$\lambda_{max}=244nm$），可能会干扰一些分光光度法的酶学分析。为获得不含 Accudenz 的相当完整的过氧化物酶体，沉淀的样品应悬于相同的溶液中，然后对轻度高渗缓冲的蔗糖（0.25 mol/L）溶液透析并更换数次缓冲液。匀浆中将匀浆器浸在冰水中以使管子保持低温。

3．过氧化物酶体纯度鉴定

按 Duve 及其同事描述的方法采用差速离心分离富含过氧化物酶体的轻的线粒体（L）组分。将该 L 组分进一步作 Accudenz 不连续密度梯度离心来分离不含其他组分的过氧化物酶体。用该方法分离的过氧化物酶体应不含任何完整的线粒体或溶酶体，但可能会有少量的内质网污染。表 10-3 描述了大鼠肝脏过氧化物酶体中的标志酶及可能污染的酶。

表 10-3　分离的大鼠肝脏过氧化物酶体中的标志酶

标志酶	定　位
过氧化物酶体的标志酶	
尿酸氧化酶	过氧化物酶体
DHAPAT	过氧化物酶体
非过氧化物酶体标志酶	
酸性磷酸酶	溶酶体污染
琥珀酸-细胞色素 C-还原酶	线粒体污染
葡萄糖-6-磷酸酶	内质网污染

四、叶绿体

所有能进行光合作用的植物都含有由叶绿素组成的细胞器，称为叶绿体（chloroplast）。叶绿体是研究得最透彻的质体，以具有称为类囊体的能进行光合作用的内膜为特征。分离和制备有活性的离体叶绿体，是研究叶绿体结构、功能、光合作用机理及其遗传控制的重要技术条件。要分离有活性、较纯的完整叶

绿体相当困难。某些植物叶细胞的细胞壁较厚，叶绿体不易从细胞中分离出来；某些植物叶绿体外膜极易破损；有些植物叶细胞中含有某些内生性抑制物质水解酶、氧化酶等，在制备过程中极易释放出来破坏叶绿体的结构和活性。因此，不是所有的高等植物都适合用来分离和制备叶绿体。

菠菜和豌豆是分离完整叶绿体的极好材料，其细胞中所含的质体较小，而且它们都能在不积累淀粉和酚类物质的条件下生长。菠菜的叶绿体/湿重比率较高，而豌豆更容易生长。此外，玉米、甜菜、大麦及低等植物如衣藻、小球藻等也是常用的材料。分离和制备叶绿体的方法很多，目前最常用且方便的是分步离心的方法。一般叶绿体的直径约为几个微米，因此细胞破碎后在一定的离心力范围内即能分离沉淀出叶绿体。这种方法不太适用于含有大而且形状不规则的叶绿体或包埋在厚厚的纤维组织中的细胞。这种情况下，可以先酶解细胞壁，然后让原生质体被动通过尼龙网使其破裂，再从中分离叶绿体[8]。后一种方法比较温和，得到完整叶绿体的产量较高。从分离的原生质体中制备叶绿体的缺点是较为费时、昂贵、且总产量较低。下面介绍以菠菜叶为材料分离和制备叶绿体的方法。

1. 操作程序

1) 取 2～3g 新鲜菠菜叶片，去叶脉后用自来水洗净，再用蒸馏水洗 1～2 次。
2) 洗净的叶片在 0～4℃冰箱中放置 1h 左右预冷（注意，为避免叶片结冻，放置时间不宜过长）。
3) 预冷后叶片剪碎，放置在玻璃研钵中（研钵置于冰浴中），加少许石英砂和少量叶绿体制备液，快速磨成匀浆，然后再加适量的制备液，磨匀后立即用六层纱布过滤，滤液移至离心管。[叶绿体制备液中含：山梨糖醇 0.33 mol/L、EDTA 2mmol/L；$MgCl_2$ 1 mmol/L、抗坏血酸 2 mmol/L、K_2HPO_4 0.5 mmol/L、$Na_2P_2O_7$ 5 m mol/L、Hepes 50 m mol/L、pH7.6（用 1 mol/L NaOH 调节）。所有的药品均需分析纯级。]
4) 用冷冻离心机（4℃，300g 离心 30s）除去完整细胞、细胞核以及石英砂等。
5) 300g 离心的上清液倒到另一离心管，在 4℃，1500g 离心 2min。叶绿体大部分在这次离心中沉淀，倾去上清液即得叶绿体沉淀。
6) 根据实验目的用一定的悬浮液（如叶绿体制备液）悬浮叶绿体。
7) 测定悬浮液中叶绿体浓度，叶绿体浓度可用叶绿素含量表示。一般叶绿体浓度调制在每毫升含 0.1mg 叶绿素为宜。叶绿素浓度测定方法如下：取 50ml 叶绿体悬浮液，溶于 20ml 80%的丙酮，离心除去沉淀，上清液在分光光度计 652nm（光径 1cm）波长下比色，测出的光密度值乘以 100/g 即为叶绿体悬浮液中叶绿素含量（mg/ml）。

2. 注意事项

1）以上操作从步骤2开始都应在低温4℃条件下进行，所用的器皿、试剂也都应在4℃冰箱或冰水中预冷。操作要求尽量迅速，温度较高和操作时间过长都会导致叶绿体完整性的破坏和活性的损失。

2）介绍的只是一种比较简单的叶绿体分离和制备的方法。不同制备方法所采用的叶绿体制备液的成分往往不同，如缓冲系统，Hepes是近几年采用的优良的缓冲系统，也可用Tris-HCl、Tricine、磷酸缓冲系统等。维持渗透压的物质除用山梨糖醇外，也可用蔗糖、甘露糖醇、NaCl等。而抗坏血酸，焦磷酸盐往往不加。根据不同实验目的有的还加含硫氢基的化合物，如巯基乙醇、二硫苏木醇等。

3）用以上介绍的方法制备的叶绿体约有60%～70%左右保持完整的外膜结构，其中也混有破碎叶绿体和其他细胞成分。因此，近几年也有人用较温和的细胞破碎方法，如以脱去细胞壁的原生质为材料分离叶绿体，这样可减少细胞破碎过程中叶绿体外膜的破坏。也有人采用密度梯度离心方法进一步纯化叶绿体。但这些方法操作繁复，操作时间过长，会降低叶绿体的活性，还有待进一步改进。

五、细胞核

细胞核由核膜包裹，包含有核基因，是基因表达的场所。为了开展细胞和遗传方面的研究工作，需要制备大量的、有一定生物活性的细胞核制剂。在大多数细胞类型中，由于核膜与内质网的连续性及细胞核表面与细胞骨架之间的连接等原因，细胞核是被固定于细胞中的。细胞核为一致密细胞器，与其他亚细胞器不同，在离心时能容易地通过浓蔗糖溶液。因此，可以分离到比较纯的细胞核。在对核蛋白质组的研究过程中，人们常用蔗糖密度梯度离心对核进行纯化[9]，但在核的纯化过程中也存在一些难题。首先，一些大的线粒体与细胞核具有相同的物理密度，常无法用密度梯度离心将它们区分而造成核蛋白的污染，同时细胞骨架蛋白与核蛋白的共纯化问题也很难避免，因为细胞核与细胞骨架是紧密相连的。为了减少细胞骨架蛋白对核的粘连，可以用氧钒核苷复合物处理匀浆后的细胞核[10]。其次，细胞核中含有大量的核酸，在蛋白的提取和分离过程中它们会产生很多问题[11]，从样品中去除核酸污染的最常用的方法是将样品同能与核酸形成复合物的底物（载体两性电解质或碱性多胺化合物如精胺）温育，再通过超速离心将这些复合物与核分离。最后也是最难解决的问题就是疏水蛋白的丢失，通常人们采用较强溶解能力的裂解液来溶解核蛋白。

完整的细胞核可以从组织或悬浮培养的细胞中分离到。肝脏来源丰富、价格

低廉、制备细胞核也较为容易，因此肝脏被广泛用来制备细胞核。这里采用的从肝脏中分离细胞核的方法是由 Blobel /Potter（1966）的方法衍生而来[12]，并根据 Cockerill / Garrard（1986）的方法加以改进[13]。这些方法都通过裂解细胞后离心来分离细胞核。

1. 操作程序

1）处死大鼠，取出肝脏，称重。另外也可以比较方便、成功地从在液氮中冷冻并保存在－80℃的肝脏中制备细胞核。使用时，将肝脏称重，放进含 3 倍体积裂解缓冲液［0.25 mol/L 蔗糖（超纯级）、10mmol/L Tris-HCl（pH7.4）、10mmol/L NaCl、3 mmol/L $MgCl_2$、1 mmol/L DTT、1.5mmol/L PMSF（用前从溶于乙醇的 100mmol/L 贮存液中添加）］的烧杯中，肝脏解冻后，继续进行下一步。

2）在一玻璃盘上，用剃刀片快速去除结缔组织，将 20g 肝脏切成约 $1cm^3$ 的小块。

3）将肝脏碎块装入含 40ml（两倍体积）裂解缓冲液的冷烧杯中。

4）将大约 30ml 这种肝脏碎块和缓冲液混合物倒入预冷的组织研磨器中，进行研磨。

5）检查细胞的裂解效率。

a. 将研磨腔放到冰上，取出 10μl 裂解液，用 1ml 裂解缓冲液稀释。

b. 取 2.7μl 稀释后的裂解液加到载玻片上，再加上 2.7μl 含 10μg/ml Hoechst 染料的裂解缓冲液，然后用一 $18mm^2$ 的 1 号盖玻片将液滴盖上。Hoechst 染料能染 DNA，可以在装有 neofluar 物镜和水银弧光灯的荧光显微镜下用蓝色滤光片观察。

c. 比较亮的 DNA 染色的细胞核与同一视野的相图。在相图中，应能看到核膜，其他细胞膜则看不到，而视野中应该看到大量细胞核。如果没有荧光显微镜，只用相差镜片也可以估计细胞裂解的百分数，未裂解细胞的细胞核是可见的，而在相差镜片下，它们明显是被包括在细胞内部。

d. 如果裂解细胞数少于 95%，则需要加大推进器速度，继续研磨样品 5～10 次。

6）在一漏斗中装入 4 层粗孔滤纸，放到一埋在冰浴中的量筒中，然后将匀浆倒进漏斗。

7）裂解剩余的肝脏碎块，匀浆液通过滤纸过滤，与其余裂解物样品混合，记录得到的裂解物体积。（用该方法 20g 的肝脏匀浆大约可以产生 50ml 的裂解物。）

8）向离心管中加入浓蔗糖溶液和裂解物（通常 7～9ml）。加裂解物时，应

将移液管头靠在离心管内壁上部，使裂解物缓慢流下，这样可减少裂解物与蔗糖溶液的混合。

9）在预冷的离心机中 23 000g，4℃离心 30min。细胞核会在离心管底部形成白色沉淀。

10）吸出上清后，每管加入 2ml 裂解缓冲液重新悬浮沉淀。

11）将沉淀的重悬浮液集中到一个 50ml 的离心管中，加裂解缓冲液至 50ml 终体积，1500g 离心 5min。

12）吸出上清，细胞核沉淀重悬浮于 1ml 的裂解缓冲液中。取出少量样品稀释 100 倍后在血球计数板上计数细胞核数目。这种制备方法一般能得到 8×10^8 或 20×10^8 个细胞核。

13）用裂解缓冲液将样品稀释到期望浓度的 2 倍，加入等体积甘油，放液氮中冷冻，－80℃保存。

2．注意事项

从肝脏中制备细胞核需要考虑的第一个问题就是用于将细胞核与其他细胞裂解液分开的蔗糖溶液的浓度：大鼠肝脏中细胞核能够通过 2 mol/L 的蔗糖溶液，而小鼠或仓鼠肝脏中得到的细胞核却不能，分离这些种类的细胞核时就要采用 1.8 mol/L 的蔗糖溶液。第二个需要考虑的问题是缓冲液的选择：制备过程中使用的缓冲液会影响分离得到的细胞核的性质。例如，用含 Mg^{2+} 的缓冲液分离的细胞核在细胞核转运实验中效果不好，而缓冲液中含有精胺和精咪会抑制细胞核在体外的裂解[14]，没有一种缓冲液适用于所有的应用，对于某一实际应用，可能需要一个“试错”的过程来找到最佳条件。第三个通常需要考虑的问题是蛋白酶解作用。PMSF 是一种传统的蛋白酶抑制剂，在使用前加到所有的溶液中；然而对于某些应用，这些步骤还不够，仍需要加入其他的蛋白酶抑制剂，例如抑蛋白酶肽、亮抑蛋白酶肽或胃酶抑素。最后需要考虑的就是一次制备过程中的肝脏用量。每个 SW28 转头（Beckman）用 20g 肝脏为最佳。这就需要 6 个超速离心管，每个装有 7－9ml 铺在约 30ml 浓蔗糖溶液上的肝脏裂解液。若在一个离心管中装更大体积的裂解液，细胞核会陷在裂解液与蔗糖溶液的交界处。如果一定要分离大量细胞核，就要进行重复操作和（或）使用更多的转头和离心机。但一定要注意，不能过大地扩大制备规模而降低处理速度，因为这样会增加蛋白酶解作用的危险性。也可以从较少量的肝脏中制备细胞核。这时就可以用较少的裂解液，进而用于加载裂解液的装有蔗糖溶液的离心管的数量也可以相应减少。如果得到的裂解液少于 7ml，就需要提高蔗糖溶液的用量，以使离心管达到 90％的填充，这样可以避免离心过程中离心管的破裂。

3. 常见问题及其通常的起因和解决方法

1）离心后沉淀中的细胞核很少。

a. 可能细胞裂解不充分。裂解效率可以用 DAPI 或 Hoechst 进行 DNA 染色并在相差和荧光显微镜下监测。如果肝脏匀浆后裂解率小于 90%，就要仔细检查肝脏是否被切得太细了。研磨器利用剪切力裂解细胞，如果肝脏碎块太小，该过程的效率就会降低。将肝脏切成 $1cm^3$ 的碎块，其细胞裂解效率最高。

b. 细胞核要通过的蔗糖溶液的浓度可能太高。如果细胞核的密度不足以通过蔗糖溶液，就会看不到细胞核沉淀。降低蔗糖溶液浓度，重复该制备过程，可以确定是不是这个原因。

c. 裂解物可能太浓，使细胞核陷在裂解物和蔗糖溶液的交界处。如果匀浆时所用的缓冲液体积少于肝脏体积的 2 倍，就会出现这种情况。

d. 离心管中加入的裂解物太多。每管不要超过 10ml，以 7～8ml 为最佳。

（2）细胞核沉淀呈红色或粉色。

向离心管中加样时裂解物与浓蔗糖溶液发生混合会使沉淀变脏。在蔗糖溶液层上面加裂解物时一定要小心。另外，离心前平衡离心管时要用裂解缓冲液，如果一定要用浓蔗糖溶液，要从离心管侧壁加入，使其沿离心管壁缓慢流下。

第二节　细胞器的组成分析

对蛋白质组组成的分析鉴定是蛋白质组学中与基因组学相对应的主要内容，它要求对蛋白质组进行表征，即所有蛋白质的分离、鉴定及其图谱化。双向凝胶电泳（2-DE）、质谱（MS）及生物信息学是蛋白质组研究的三大技术支持。2-DE用来分离复杂样品中的蛋白质；蛋白质的肽质量指纹谱（PMF）和蛋白质数据库的搜索常被用来鉴定从 2-DE 上分离出的蛋白质。但是，由于真核细胞的复杂性，想一步鉴定一个细胞或组织中的全部蛋白质几乎是不可能的。因为真核细胞的复杂性不仅只在于具有低丰度表达的蛋白质，还包括具有极酸、极碱等电点的蛋白质；疏水蛋白；极大、极小相对分子质量等蛋白质的存在。这些蛋白质极易在常规的 2-DE 中丢失。为了达到分析一个组织或细胞中的全部蛋白质的目的，现在人们更倾向于亚蛋白质组如特定的细胞器（细胞膜、细胞核）、细胞中大分子结构等的研究。

亚细胞的分离或细胞器的纯化，为蛋白质组分析中样品制备提供了一个有效的方法。这种方法的优点在于与细胞总蛋白相比，用细胞器蛋白作为分析对象，降低了样品复杂性，而且细胞的功能单位如细胞器、大分子结构或多蛋白复合体

也没有被破坏。从大量的研究中可以看出，样品的复杂性越低，研究的成果就越详细、越有说服力。特定细胞器的组成分析可以先进行该细胞器的分离，然后将提得的细胞器溶于 60～100μl 的样品裂解液中［7mol/L urea，2mol/L thiourea，1% DTT，2% CHAPS，0.8% Pharmalyte (pH=3～10)］，之后以 12 000 g 离心 30min，取上清用于二维电泳的分离分析用，剩余样品储存于 -70℃[15]。当然这种方法也有其缺点，如为了得到足够的细胞器蛋白用于分析，需要比较复杂的样品制备过程，而且它们的纯度很难鉴定，可能会有其他细胞器污染的问题存在。因此，为了得到某个感兴趣的蛋白的准确细胞定位，还需要补充其他的实验，如免疫电镜实验或免疫荧光实验[19,20]。目的蛋白与标签蛋白的过表达[21]（如 myc-tag）或与绿色荧光蛋白的融合蛋白[22]在细胞中的定位分析也可以作为蛋白定位的一项指标。如果蛋白过表达的结果仍不能令人满意，还可以用 TAP 标签的方法加以验证[23]。此外，还可以建立一个感兴趣蛋白的缺失体，通过该蛋白功能的间接证据，也可以推断该蛋白的细胞定位。但是缺失体的建立有赖于可检测表型的存在。

我们相信将来生物信息学的工具，如预测蛋白质亚细胞定位的程序将为蛋白质组的研究提供很有价值的提示。用计算机预测蛋白质的亚细胞定位将更有利于高通量检测蛋白的需要，并将成为新蛋白筛选的有用工具。目前生物信息学中有关特定细胞器蛋白质组的数据库还正在建立，处于刚刚起步的阶段。

SWISS-PROT 蛋白数据库现在包括 269 个人线粒体蛋白。专门的数据库，如线粒体蛋白质组数据库 MITOP[16]也已经建立，该数据库全面地包含已经鉴定了的线粒体蛋白质组，特别是与人类线粒体疾病相关的蛋白。目前该数据库共包含有 225 个蛋白。Rabilloud 及其同事已经建立了人胎盘中线粒体蛋白的二维电泳参考图谱，该图谱已经公布于网站(http://www-dsv.cea.fr/thema/Mitopick/Mito2D.html)，其中有 53 个蛋白质已经通过肽质量指纹谱、N 端测序或免疫杂交得以鉴定。目前这一研究还没有发现任何新的线粒体蛋白，因为大部分鉴定的蛋白已经存在于 SWISS-PROT 数据库中了。在 SWISS-PROT 数据库中，有 157 个蛋白质定位于高尔基体中，Taylor 用蛋白质组技术分析了大鼠肝细胞的高尔基体组成[17]，其中包含 173 个蛋白质点；SWISS-PROT 数据库中的 35 个蛋白被定位于过氧化物酶体，目前还没有纯化后的过氧化物酶体蛋白质组数据库被建立。

现在已有 946 个人的核蛋白存在于 SWISS-PROT 数据库中，足以显示科研工作者对细胞核蛋白的重视。细胞核蛋白专门的数据库已经在 1993 年建立[18]。2000 年已经建立了一个专门的核蛋白数据库，并公布于网站(http:// www.hgu.mrc.ac.uk/ Users/ Wendy.Bickmore/ npdintro.html)。为了研究与凋亡有关的蛋白质，BL60(Burkitt Lymphoma)细胞系的核蛋白的二维电泳参考图谱已经建立，其中有 32 个被建立，主要属于核基质蛋白。关于用蛋白质组技术建立核蛋白质组，可以参考人肝的 SWISS-2DPAGE 图谱，该图谱可以通过 EXPASY 的服务器得

到(http:// www.expasy.ch/ ch2d/)。

建立细胞器的参考图谱将会对研究特定细胞器的科研人员提供有用的工具。这种二维电泳参考图谱将细胞蛋白进行了初步定位，与细胞总蛋白相比具有更强的实用性，它可以帮助我们发现某个细胞器中特有的新的通路，此外，可以帮助我们更为详细地了解由于某个细胞器中不正常蛋白的表达而导致的疾病。目前亚细胞器研究中发现的新蛋白还较少。

到目前为止，还没有对已鉴定蛋白的可能的翻译后修饰建立数据库。蛋白的翻译后修饰不仅对蛋白的生理功能起到很大作用，而且这种生理功能很难从蛋白的氨基酸序列上估计出来。所以我们相信在将来的蛋白质组分析中蛋白质翻译后修饰的研究将会逐步受到重视。现在许多研究中，将液相色谱与普通的 SDS-PAGE 结合代替了单独使用 2-DE 来分离蛋白的方法，尽管这种方法避开了难溶蛋白质的问题，但 2-DE 同时还会提供等电点及分子质量的信息，这些信息对于蛋白的鉴定十分重要，更为重要的是，2-DE 会将同一蛋白质的不同存在形式分离开，这对于研究蛋白质翻译后修饰有重要意义。随着蛋白质组分析技术的不断提高，我们相信整个细胞及组织的蛋白质组组成的分析终会实现，而将蛋白质组的研究与细胞生理研究相联系则会使我们更为详尽地了解细胞的功能。

参 考 文 献

1 Bérengère P B, Florence B, Hartmut B, Ernst W M, Jose A G S. Translation control : bridging the gap between genomics and proteomics? TRANS in Biochemical Sciences, 2001, 26(4): 225－229

2 Richwood D. Preparative centrifugation: A practical approach. Oxford, 1992

3 de Duve C. Tissue fractionation-past and present. J. Cell Biol, 1971, 50: 20d－55d

4 de Duve C, Beaufay H. A short history of tissue fractionation. J Cell Biol, 1981, 91: 293－299s

5 David L S, Robert D G, Leslie A L, Cells: A Laboratory Manual. Cold Spring Harbor Laboratory, 1998

6 Morré D J. Isolation of Golgi apparatus. Methods in Enzymol, 1971, 22: 130－148

7 Kindl H, Lazarow P B eds. Peroxisomes and glyoxysomes, vol 236. Annuals of the New York Academy of Science. New York Academy of Science. 1982

8 Somerville C R, Somerville S C, Ogren W L. Isolation of photosynthetically active protoplasts and chloroplasts from *Arabidopsis thaliana*. Plant Sci Lett, 1981, 21: 89－96

9 Eva J, Christine H, Diego C, Jean-Charles S, DenisF H. The establishment of a human liver nuclei two-dimensional electrophoresis reference map. Electrophoresis, 2000, 21: 3483－3487

10 Pasquali C, Fialka I, Huber L A. Subcellular fractionation, electromigration analysis and mapping of organelles. J Chromatogr B, 1999, 722: 89～102

11 Rabilloud T. Solubilization of proteins for electrophoretic analyses. Electrophoresis, 1996, 17: 813－829

12 Blobel G, Potter V R. Nuclei from rat liver: Isolation method that combines purity with high yield. Sciences, 1966, 154: 1662－1665

13 Cockerill P N, Garrard W T. Chromosomal loop anchorage of the Kappa immunoglobin gene occurs next to the enhancer in a region containing topoisomerase Ⅱ sites. Cell, 1986, 44: 273～282

14 Newsport J, Spann T. Disassembly of the nucleus in mitotic extracts: Membrane vesicularization, lamin disassembly, and chromosome condensation are independent processes. Cell, 1987, 48: 219～230

15 Bollag D M, Edelstein S J. Protein Methods. Chapter 2: protein extraction, Wiley-Liss. NewYork, 1991

16 Scharfe C, Zaccaria P, Hoertnagel K, Jaksch M, Klopstock T, Dembowski M, Lill R, Prokisch H, Gerbitz K D, Neupert W, Mewes H W, Meitinger T . 2000. MITOP, the mitochondrial proteome database: update. Nucleic Acids Res, 2000, 28: 155～158

17 Taylor R S, Fialka I, Jones S M, Huber L A, Howell K E. Two-dimensional mapping of the endogenous proteins of the rat hepatocyte Golgi complex cleared of proteins in transit. Electrophoresis, 1997, 18: 2601～2612

18 Wirth P J, Luo L D, Benjamin T, Hoang T N, Olson A D, Parmelee D C. The rat liver epithelial (RLE) cell nuclear protein database. Electrophoresis, 1993, 14: 1199～1215

19 Wigge P A, Jensen O N, Holmes S, Soues S, Mann M, Kilmatin J V. Analysis of the Saccharomyces spindle pole by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. Cell Biol, 1998, 141: 967～977

20 Rout M P, Aitchison J D, Suprapto A, Hjertaas K, Zhao Y, Chait B T. The yeast nuclear pore complex: composition, architecture, and transport mechanism. J. Cell Biol, 2000, 148: 635～651

21 Fialka I, Steinlein P, Ahorn H, Bock G, Burbelo P D, Haberfellner M, Lottspeich F, Paiha K, Pasquali C, Huber L A. Identification of syntenin as a protein of the apical early endocytic compartment in Madin-Darby canine kidney cells. J Biol Chem, 1999, 274: 26233～26239

22 Neubauer G, King A, Rappsilber J, Calvio C, Watson M, Ajuh P, Sleeman J, Lamond A, Mann M. Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. Nature Genet, 1998, 20: 46～50

23 Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Searphin B. A generic protein purification method for protein complex characterization and proteome exploration. Nature Biotechnol, 1999, 17: 1030～1032

第三节　蛋白复合体分离及组成分析

蛋白质是细胞生物功能最主要的执行者。几乎一切生命现象都要通过蛋白质的结构与功能体现出来。有效功能的实现绝不是单一成分能够承担的，而是众多蛋白分子共同参与、作用的综合。这些蛋白分子往往作为其中的一个亚单位，各成员之间彼此相互作用而聚集成大的蛋白质功能单元，共同完成某种特殊的“使命”，通常将这样的大蛋白分子的集合称作“蛋白复合体”或“蛋白机器”(protein assembly，protein machine)。细胞作为蛋白复合体的集合，犹如一个奇妙的工厂，依靠这些蛋白与蛋白之间精细、完备的相互作用，形成了复杂、有序的生命活动网络，从而使机体能够适应千变万化的外部世界。这些组合遵循严格的时空控制，具有高度协调一致的活动组分，它们相互依赖，相互影响，相互制约，共同执行某个或某一类生物功能，而这些大分子复合物发挥作用的环节往往是复杂生命网络中重要的检查点或重要的调控位点。作为后基因组时代蛋白质组计划的重要课题，世界上越来越多的研究小组致力于蛋白相互作用研究，尤其是对蛋白复合体结构与功能的研究热潮不断高涨。具有代表性的是 1998 年生命科学权威杂志 CELL 第 92 卷以一系列综述的形式将蛋白复合体确立为一个崭新的研究

领域，目的在于探讨各种复合体结构组成、亚基成分之间的相互作用以及它们在相应重要生化事件中的调控机制，进而揭示蛋白质相互作用的一般规律，逐步了解细胞内复杂调控网络的概貌，力争发现有价值的作用形式和其中重要的蛋白参与分子，以确定与重大生命活动（胚胎发育、重大疾病）相关的蛋白质组基础。

根据现有文献报道，目前比较明确、有一定功能提示的蛋白复合体包括：

1）参与基因表达（转录、翻译）调控的复合体：拓扑异构酶（topoisomerase）、多聚酶（polymerase）、拼接体（spliceosome）、复制体（replisome）、IKK 体；

2）参与蛋白质合成反应的复合体：核糖体（ribosome）、内质网蛋白质合成复合体（ER protein synthesis complex）；

3）调节蛋白降解代谢的复合体：泛素-26S 蛋白体复合体；

4）参与细胞凋亡、启动 caspase 级联的复合体：凋亡复合体 apoptosome；

5）介导细胞周期调控的周期复合体：cyclosome /APC、SCF complex；

6）错误折叠蛋白形成的聚集体：aggresome；

7）信号转导复合体：COP9 / JAB1 signalosome；

8）其他类型的复合体：如 ATP 合成复合体（ATP synthesis complex）、中心体（centrosome）、SWI/SNF complex、膜受体复合物、细胞骨架蛋白、分子伴侣复合物等。

与独立的蛋白质作用模式类似，众多的大蛋白质复合体也具有高度的“社会性”，它们在生命过程的各个环节中几乎无处不在，既体现了责任明确，分工负责；又充分保证了整体上的协调一致与高度统一。蛋白复合体之间的确有错综复杂的作用网络存在，比较典型的是由若干复合体所构成的一条功能连续的“主线”——即正常情况下调节细胞周期的 cyclosome/APC 与信号转导的 signalosome、以及特殊状态下参与凋亡过程的 apoptosome，它们的共同下游事件是将所需加工、处理、降解的蛋白分子交于泛素依赖的蛋白酶复合体来执行，它是这些重要生理步骤的核心环节。一旦这一共有通路被阻断，就会导致重要生物事件的停滞、靶蛋白的累积（如，由错误折叠蛋白聚集而形成的 aggresome），从而引发疾病（如疯牛病等）。本节将着重就细胞内几个具有重要生物学功能的蛋白复合体的研究方法进行介绍。

一、拼接复合体研究方法

（一）概述

早在 1985 年，Grabowski 和 Brody 等[1,2]在进行体外剪接反应时，通过甘油密度梯度沉降（60S）鉴定并命名了拼接复合体。人们发现 RNA 的剪切系统与核糖体一样巨大。拼接复合体的“组成列表”中包括 5 个小核 RNA（snRNA）

及 50 多种性质稳定的蛋白质。因此，拼接复合体与核糖体一样被称作核糖核蛋白器。它们都是耗能机器，但都不需要水解 NTP 进行基本化学反应，因为催化步骤是通过两次转酯反应实现的。尽管装配的早期阶段是非能量依赖的，但所有的后续装配步骤都是依赖于 NTP 的水解。与这些步骤相吻合的是：RNA 螺旋的形成与崩解，每一个 ATP 依赖的步骤需要一个或多个 ATP 酶超家族的蛋白质参与，它们与 DNA 螺旋酶具有序列相似性。

对于由 spliceosome 所驱动的基本化学反应，可以由 RNA 单独催化。在观察真核蛋白质合成机器时发现，这一过程需要耗能驱动分子动点、时钟和启动源。由于拼接必须有效和准确，我们估计，spliceosome 同样也要利用上述装置，以达到高速和准确。对于这种相似性有一种普遍的但未经证明的假说：拼接体成员通过消耗 ATP 解开 RNA。新近证实，拼接体“DexD/H box” ATP 酶作为一步反应的动点解开短的 RNA 双倍体，RNA 重排的精确性表明这些 RNA 依赖的 ATP 酶还可作为时钟监测拼接的步骤。

（二）研究方法

1．RNA 前体合成与体外拼接活性实验

以帽子起始方式，在转录同时掺入 ^{32}P 作为示踪分子，在体外加入酵母或 HeLa 细胞核提取物、代谢能量（肌苷、ATP）以及离子缓冲液等必需成分，实现拼接复合体在体外的重组[1,2]。

2．复合体提取、纯化与鉴定

主要研究对象为模式生物——酵母、果蝇；哺乳动物研究主要集中于细胞水平（以 HeLa 细胞研究最为多见），未见有从组织中直接提取拼接复合体的文献报道。主要方法为甘油密度梯度离心，也可以用蔗糖密度梯度离心（Frendewey 1985）来实现分离。Grabowski 在 1985 年建立的提取方法至今已发展成为相当成熟的分离、纯化方案：在含有 10%～30% 的连续甘油梯度中加入约 125μl 的 pre-mRNA，以 40 000 r/min 超速离心 3h，分成约 20 个 560μl 的组分；从每一组分中取出 20μl 样品，进行 ^{32}P-cpm Cerenkov 梯度谱活性分析；并取出奇数梯度组分经纯化后，进行 PAGE 电泳分离和 Northern 杂交分析[1]。

3．生化特性、组成成分与相互作用分析

对于分离出的复合体成分可经亲和纯化后，采用 2-DE 分离（Shevchenko et al, 1996）、质谱鉴定以及数据库检索，了解其蛋白组成与一般生化特性及功能线索[3]。主要采用如下技术方法：

(1) 酵母系统与免疫共沉淀相结合

这两种目前研究蛋白之间相互作用的最主要技术，同样也在分析 snRNP 和参与拼接的附属蛋白所形成复合物的组成、以及验证它们之间的相互作用中得到广泛的采用。Awasthi 等研究发现，哺乳动物 70kDa 蛋白作为 U1 snRNP 的成分之一，与一系列蛋白相结合，通过双杂交获得 4 个 U1-Snp1p 相关（Usa）蛋白；进一步免疫沉淀实验验证了其中的一个蛋白为已知的拼接复合体关键成分——Prp8p。这种作用反映了 Snp1p 及 U1 snRNP 在拼接复合体形成晚期的一些新的功能和在调节 mRNA 前体（pre-mRNA）的组成性和选择性拼接方面的重要作用[4]。Ambrozkova 等利用双杂交获得与拼接因子小亚基 U2AF（spU2AF23）C 端相互作用的裂殖酵母共调节分子 spSNW1，与人类共调节分子 SNW1/SKIP/NcoA-62 直系同源，将这两个相互作用的蛋白加上标签后在裂殖酵母中进行表达，用核提取物成分进行 CO-IP 实验也验证了上述结合作用。提示了这些转录共调节分子在拼接复合体的招募中发挥作用[5]。

(2) 免疫封闭（immunodepletion）

这是分析蛋白成分功能作用的一个较好的方法。Makarova 在分析 U4/U6.U5 三元- snRNP 复合物（人类有两种：65kDa 和 110kDa）与前拼接复合体的相互关系时，封闭 HeLa 细胞核提取物中的这两种蛋白，可以阻断成熟拼接复合体的形成，并进一步抑制了 mRNA 前体（pre-mRNA）的拼接，但不影响 U4/U6.U5 三元- snRNP 复合物的完整性。在加入重组 65kDa 和 110kDa 蛋白后，拼接复合体的组装和拼接催化活性得到恢复。说明这两种蛋白在将 U4/U6.U5 三元- snRNP 复合物招募到拼接复合体前体上起关键作用，但不参与三元- snRNP 复合物稳定性的维持[6]。Ajuh 也用同样的方案，利用抗 CDC5L 抗体，在体外阻断了 mRNA 前体拼接产物的形成；加入在 HeLa 细胞核提取物中纯化的 CDC5L 复合物后，又恢复了 mRNA 前体拼接活性[3]。

(3) 构建蛋白的缺失或突变体

通过基因剔除及传统的分子克隆技术构建目的蛋白成分的缺失或突变体，可在体研究该蛋白在 mRNA 前体拼接过程中所起作用的大小、作用的时相及发生缺陷时所导致的后果。Blencowe 小组采用这种特定的缺失方式研究了两种拼接共激活子 SRm160/300 的作用，结果表明，缺失 SRm300 并不阻断 mRNA 前体拼接；而 SRm160/300 联合缺失将严重影响拼接活性；但单独加入重组 SRm160 可以显著提高拼接活性。这一结果清晰地说明 SRm160 可能在 SRm160/300 共激活子中作用更为关键，从而澄清了这两种核基质抗原在促进拼接因子与 mRNA 前体的结合这一重要步骤中起同等作用的认识[7]。

4. 质谱技术

Labourier 运用该方法结合多肽片段测序和数据库检索，确定果蝇拼接复合

体早期成分 U1 snRNP 中的整合分子为 70K、SNF、B、U1-C、D1、D2、D3、E、F 和 G 蛋白。U1 -相关内在激酶活性从人到果蝇高度保守，体外实验中具有一致的底物特异性[8]。另外，Ajuh 通过质谱技术鉴定了首次纯化并具一定功能提示的、包含 CDC5L 人类非 sn 拼接复合物亚基中的主要蛋白成分以及至少 5 种辅助蛋白因子[3]。可见，质谱技术的确是蛋白相互作用研究中不可或缺的关键手段。

5．三维结构分析

通过对 snRNP 的蛋白-蛋白复合物、RNA -蛋白复合物的 X 射线衍射晶体结构图谱分析，可以直接观察出 mRNA 前体拼接的详细分子机制、了解 snRNP 的总体构成全貌、以及蛋白-蛋白、RNA -蛋白之间相互识别的作用机制[9]。

此外，Newby 利用液相 NMR 光谱检测酵母中 U2 snRNA 与内含子序列配对形成的二聚体，来了解分支点促进识别和化学活性的结构特征。通过在保守区域掺入合成的假尿嘧啶寡聚物，导致螺旋参数的显著偏移，以及对未配对的腺苷酸产生螺旋外定向。由此改变了保守区 U2 snRNA 的结构、提高了分支点序列的稳定性；分支点腺苷酸的螺旋外定向促进了拼接过程中对这一碱基的识别[10]。

二、蛋白酶复合体研究方法

（一）概述

对于蛋白酶复合体，其实我们并不陌生，生物化学的研究领域中很少有像蛋白水解酶这样成功地开发应用于临床医学及产业化生产的。早期的该类生物工程产品是微生物和植物的蛋白酶作为去垢剂和肉类保鲜剂，而最重要的贡献是利用 DNA 重组工业生产出组织纤溶酶原激活剂，从而为心肌梗塞的治疗掀开了革命性的一页。近年来涌现出的若干最有价值的药物是能够选择性地抑制 HIV 蛋白酶以及血管紧张素转换酶的蛋白酶抑制剂。在这些成功经验的基础上，许多制药公司现已着手研究开发对于血液凝集、炎症、肿瘤转移、病毒包装等病理过程相当关键的蛋白酶的抑制剂，以期获得相关疾病有效治疗的新途径。

真核细胞内蛋白质的降解是多步骤的复杂过程，由一个巨大的蛋白水解复合物——26S 蛋白酶体来执行，并依赖共作用因子——泛素提供高度选择与精确调节。在哺乳动物细胞内，全部蛋白在不断进行着合成和分解，但对于蛋白质个体而言，其半衰期从几分钟至几天不等。这一降解过程具有高度选择性及精密的调节。它最关键的功能之一就是迅速消除调节蛋白，例如控制细胞周期转换各种步骤的细胞周期蛋白[11,12]。而且，在不同的生理条件下，蛋白质的降解速率发生非常显著的变化。由此，蛋白水解可作为时钟装置通过水解关键蛋白来调整基因表达的速度并经生化途径及其他细胞过程等予以输出[11~13]。

细胞内蛋白水解的另一关键作用在于作为质量控制装置，选择性地破坏那些由于突变而不能正确折叠的蛋白或合成后损伤的蛋白。真核细胞依赖蛋白酶系统将预降解的蛋白加上底物标签（泛素化），由大的蛋白复合体——26S 蛋白酶复合体的核心—20S 亚复合物来执行水解功能，这一途径是经过严格调控的。26S 蛋白酶复合体相对分子质量为 2.4MDa，含有 30 多种不同的亚基[14]，比典型的消化蛋白酶大 100 倍，而后者的作用方式如“一对剪刀”，将蛋白多肽断裂为两个片段。与之相反，20S 的蛋白酶复合体（26S 蛋白酶复合体的活性核心）是一个 4 个环形成的中空的隧道状结构，包含多蛋白酶解位点，如“绞肉机般”将非需要蛋白降解成碎片。真核蛋白酶复合体具有断裂多种肽键的能力；在其 6 个活性位点中，2 个具有胰凝乳样特异性，2 个具有酪蛋白样活性，2 个为 caspase 样活性。新近的研究表明：活性位点之间可以互相控制，而且在降解蛋白的过程中以相当有序的形式发挥着功能[15]。

结构组分（表 10-4）

表 10-4

Protein machine	species	MW	constitute	shape	function
	Enkaryotes	2.5MDa	20S	barrel shaped/cylinder-shaped	proteolytic
26S					core complex
Proteasome				10-11 * 15nm	degrade Ub-protein to oligopeptides
(31submits)			19S	cap at one or both ends of 20S	regulatory
			(PA700)	complex	
			11S	activator	
20S	Archaent		α, β		
Proteasome	Bacteria		(14 copies of each submits)		

真核细胞 20S 蛋白酶复合体由 14 种不同的亚基构成，可以分为 α、β 两个亚类，每类亚基各含有两个拷贝，形成排列为 $\alpha_7\beta_7\beta_7\alpha_7$ 形式的 4 个七次跨膜环，即两个 α-外环和两个 β-内环，所有的水解活性被认为存在于 3 个 7β 类亚基中。这些亚基共同组成三个内腔（5 nm in φ）：由相邻的两个 β-环组成一个中央腔；由一个 α-环和一个 β-环组成两个外腔。多以 β 类型亚基的两个组分 LMP2 和

LMP7 作为标记蛋白，检测蛋白酶复合体的分布、定位及含量。α、β类型亚基主要区别：α类亚基高度保守的 N 端延伸部分（20～30 残基）形成，α螺旋（HO）跨过中央 βsandwich 顶端，其功能目前不十分清楚，但由于它位于 α 环的顶端，靠近前室的入口，表明它可能对底物的转位及蛋白酶复合体和其调节复合物的相互作用具有重要意义[14,16,17]。

（二）研究方法

1．提取、纯化蛋白酶复合体

（1）直接法

目前，蛋白酶复合体研究一个重要的方面在于探讨癌症晚期恶液质状态、未经治疗的糖尿病患者、艾滋病患者消瘦、肌萎缩的机理。已有证据表明，蛋白质降解骨骼肌的分解代谢过程与蛋白酶复合体活性增高相关[18,19]。因此，从不同种属来源的骨骼肌组织中分离提取蛋白酶复合体的文献比较多见（详见表 10-4）。以大鼠为实验材料的研究报道也很多[20～23]。Farout 等从大鼠的 10 种组织的粗提成分中分析了 5 种蛋白酶活性，发现个别蛋白酶活性的组织特异性差异反映了复合体组成的组织特异性或其调节成分的差异，但并不代表蛋白酶复合体在不同组织的分布存在差异[20]。Piccinini 也从牛脑中获取了 20S 和 26S 的蛋白酶复合体，并对其结构与功能进行了描述[24]。简要来说，分离的程序包括：组织匀浆、离子交换色谱、HPLC/FPLC、凝胶过滤/Mono Q 层析法、甘油梯度超速离心等。

（2）间接法

由于在体分离复合体的步骤相对比较繁琐，有人则单纯以 20S 的代表亚基成分 LMP2 和 LMP7 作为检测指标反映蛋白酶水平的改变情况，来分析附加因素与蛋白酶复合体功能的关系。如 Nelson 等采用免疫印迹法和免疫组化法检测了内毒素（LPS）作用下 3d 后，大鼠组织（心、肾、肺、脑、骨骼肌）中 LMP7 的表达水平，发现 LMP7 的表达均有不同程度的上调，提示机体在对感染和炎症产生反应时所造成的器官功能失调，导致了蛋白酶复合体活性的增高，这与在紧急条件下（营养不良、创伤、患病）蛋白质的分解代谢途径活跃起来，提供氨基酸转化为葡萄糖以提供能量相吻合[25]。

2．分离、鉴定蛋白酶复合体组分

纯化后的蛋白酶复合体中的组分常通过 2-DE、质谱或两者相结合来进行鉴定。一个典型的例子是 Verma 等采用大蛋白复合体组成直接分析法（DALPC），通过质谱技术，鉴定了超过 24 个蛋白复合体中相互作用的蛋白质，其中包括出芽酵母 19S 帽子结构的一个新亚基 Ylr21c（Daq1）、与哺乳动物 19S 帽子结构

(PA700) 同源的亚基 Ygr232w (Nas6)、以及热激蛋白 Hsp70、Hsp82、脱泛素化酶 Ubp6、转录调控蛋白、有丝分裂蛋白、微管组装蛋白、RNA 代谢蛋白、信号转导蛋白等[17]。

3. 蛋白酶复合体活性分析

20S 蛋白酶复合体具有 5 种不同的肽酶活性：胰凝乳蛋白酶样 (Cht-l)，胰蛋白酶样 (T-L)，肽酰谷氨酰氨肽水解 (PGPH) 活性、支链氨基酸优选 (BrAAP) 及小中性氨基酸优选 (SNAAP) 活性，分别在基本羰基基团、疏水性、酸性、支链、中性氨基酸部位水解合成肽。在提取的蛋白酶复合体成分中加入相应的荧光寡肽底物，利用荧光分光光度计检测水解后底物肽片段的荧光值，反映出酶活性水平的高低。另外，酪蛋白也是一个广泛应用的蛋白酶复合体酶解活性分析底物，而且它在不同组织中的活性水平差异不大。因而，以 [^{14}C] -甲基酪蛋白为底物经复合体降解后，检测三氯乙酸可溶的放射同位素释放量，灵敏性更高，结合其他酶解分析，使结果更加可靠[20]。

表 10-5 蛋白酶复合体的组织来源

Species \ Organ	Muscle	Liver	thymus	brain	pituitary	epidermis	Skin	placenta
Human						+		+
Rat	+	+	+				+	
Bovine				+	+			
Porcine	+							
Rabbit	+							
Fish	+							
Ovine	+							
Lobster	+							

三、聚集体研究方法

(一) 概述

细胞通过分子伴侣和蛋白酶复合体降解错误折叠的蛋白质等抑制集聚形成等机制，避免潜在毒性聚集物的累积，这种集聚一旦形成，往往可以抵抗蛋白水解并集聚于包含体内。最近的研究表明，在动物细胞中，集聚蛋白通过达因蛋白依赖的沿微管逆向转运的特异转运机制到达包含体。这种微管依赖的包含体排列在

核旁或核周区域（juxtanuclear/perinuclear），被含波形纤维蛋白（vimentin）的中间纤维所包绕，称作 aggresome（1988 提出）。aggresome 由可溶的、非泛素化蛋白和整合跨膜蛋白泛素化部分来形成，并伴随着中间纤维蛋白 vimentin 的再分布，形成一个围绕未降解聚集物的核中心粒中心的笼状结构。aggresome 在核周的聚集干扰正确的 Golgi 定位和微管的正常星状分布，但是，ER-to-Golgi 蛋白转运通常发生于含有 aggresome 的细胞中，从而支持了 aggresome 形成可能是细胞针对错误折叠蛋白存在的一种普遍反应这一假说[26~28]。

（二）功能与意义

aggresome 的形成是易于集聚的错误折叠蛋白的产生超过蛋白酶的降解能力时，细胞所作出的一种普遍反应。蛋白质在细胞和细胞外的聚集往往与许多疾病相关，特别是错误折叠的蛋白质在含有泛素的细胞质包含体中聚集，参与了包括老年神经退行性病变以及系统性淀粉样变性等多种疾病的发病过程。尤其要注意那些维持细胞蛋白正确的翻译后转运的残基，如果一旦目标蛋白的这些残基发生突变而错误折叠，引发错误的靶向性，就会形成稳定的 aggresome 复合体，最终导致疾病的发生。

（三）研究方法

1．定位检测

采用免疫荧光和 transmission electron microscopy（透射电镜术）方法证实 aggresome 的存在；在核旁或核周区域的 aggresome，可以经泛素（ubiquitin）染色定位，并含有微管蛋白（tubulin）和 vimentin。

2．蛋白酶复合体抑制实验

抑制蛋白酶活性将平衡转向 aggresome 的形成；提高细胞内分子伴侣 Hsp70 的表达可以减小细胞内的集聚并定向于蛋白酶复合体。转导 Hsp70 的靶蛋白进入细胞后，同样降低了 aggresome 的形成，增加目标蛋白的降解。共转染 GFP－目的蛋白片段（具有 myc 标签的泛素）以及与具有特定抗蛋白酶复合体抗体的共定位实验表明：该片段可以招募外源性泛素和蛋白酶复合体于新合成的 aggresome 中。在转染的 293 细胞和（或）CHO 细胞中，过度表达目的蛋白或应用蛋白酶复合体活性抑制剂，将导致稳定的、高分子质量的、去垢剂不溶的多泛素化、未经降解的目的靶蛋白的聚集。上述结果体现了 aggresome 与蛋白酶复合体之间的联系[29~30]。

3．与细胞骨架的关系

采用诺考达唑（nocodazole，抗肿瘤药）崩解微管，可完全中止目的蛋白片

段在一个 aggresome 内的集聚，引起这些结构扩散至细胞周围。突变目的蛋白，结合蛋白酶复合体抑制剂 lacatarysteine，会增加 aggresome 形成，这一过程可以被 nocodazole 共孵育所阻断。体内（*in vivo*）time -lapse 分析 aggresome 动力学表明：在细胞周围的微小聚集成分沿着微管向 MTOC 部位运转。过度表达细胞蛋白嵌合体 GFP-250 可促进 aggresome 形成，同时伴有中间纤维蛋白 vimentin 的再分布，以及蛋白酶复合体及 Hsp70 和分子伴侣家族的伴侣蛋白质的招募。P50/dynamitin 的过度表达可以促使达因动蛋白复合体的解离，显著地抑制 aggresome 的形成，表明细胞质内“微末端指导的马达活性”（minus-end-directed motor activities）是 aggresome 形成所必需的[30,31]。

四、凋亡复合体研究方法

（一）概述

凋亡这一进化上保守的过程对于机体在发育过程中胚胎形成、器官形态改变、去除非需要细胞、维持组织的稳态具有十分关键的作用，一旦失调，将会导致多种疾病。凋亡所引起的细胞死亡是以程序性和执行性为特征的，通过激活一组半胱氨酸/精氨酸蛋白酶（caspases）引发一系列典型的形态学变化和生化改变。细胞中 caspases 的组成型表达通常是以非活性的酶原形式，需经蛋白水解切割后激活。目前已鉴定出 14 种 caspases，可将它们分成两大类：“起始（initiator）caspases”，如 caspase 8、9、10，具有长的前体结构域；可以直接或间接激活的“效应（effector）caspases”，如 caspase 3、6、7，含有短的前体结构域；“起始（initiator）caspases”的激活是通过其长的前体结构域与 caspases 招募结构域（CARD）或接头分子的死亡效应子结构域（DEDs）发生相互作用而形成的分子复合体。因此，细胞内存在两种不同的凋亡信号通路：一是通过 caspase 8/10 控制的 CD95（Fas/APO-1）受体介导的细胞死亡，形成死亡诱导信号复合体（DSIC）；二是死亡信号刺激线粒体释放 cytochrome c（Cyt-c）并激活 caspase 9/Apaf-1/ cytochrome c 凋亡复合体，从而引发一系列 caspases 级联反应，最终将细胞引向死亡[32~34]。

凋亡复合体（apoptosome）是一种含有 Apaf-1（apoptotic protease-activating factor -1）、在 ATP/dATP 和 Cyt-c 存在下催化前 caspases 激活、经体外重组形成的约 1.4MDa 的大分子复合物；而体内内源性的活性 apoptosome 则为约 700kDa 的复合物[32]。作为线粒体凋亡途径的一个关键调节子，Apaf-1 是首先在哺乳动物中被发现的接头分子，分子质量约 130 kDa，其在线虫中的同源物为 CED-4。它由三种功能域组成：①N 端 caspase 招募结构域（CARD）可与前 caspase 9 相结合；②CED-4 样区域可实现自身的寡聚化；③C 端具有 WD-40 重复序列的调节区，遮盖着 CARD 和 CED-4 部位。在凋亡过程中，Cyt-c 和 dATP 消除了

WD-40重复序列的抑制作用，使 Apaf-1 可以发生寡聚化，随即招募并激活前 caspase 9。Lauber 等发现不同的刺激信号诱导 caspase 介导的 Apaf-1 裂解，断裂位点发生在 N 端信号肽除去 CARD H1 螺旋所形成的 84 kDa 的片段。另外的断裂点位于 WD-40 重复序列，使 p84 在 cytochrome c 缺乏的条件下也可以发生自身寡聚化，形成约 440 kDa Apaf-1 多聚体。由于 CARD 的部分缺失，导致 p84 多聚体缺少 caspase 9 或其他 caspase 的活性。因而，Apaf-1 的断裂可能是加工后的 caspase 从 apoptosome 释放出来的原因[35]。

1．apoptosome 与 caspase 级联

在外界凋亡信号的诱导下，线粒体膜通透性的改变导致 Cyt-c 释放至细胞浆，结合并激活 Apaf-1，从而使线粒体外损伤与 caspase 激活机制偶联起来。Cyt-c 与 dATP 或 ATP 结合，通过 NB-ARC 结构域形成一个大的 Apaf-1 寡聚体（推测为一个十聚体）；同时暴露 CARD 结构域，与 caspase 9 前体的 N 端前结构域的 CARD 互补区相互作用，通过引发靠近效应，邻近的前 caspase 9 酶原相互发生水解，产生典型的激活 caspase 群的大小亚基[36～42]。活性 caspase 9 结合于 Apaf-1 寡聚体，直接断裂并激活前 caspase3，后者是一个效应子，断裂并激活其余的下游 caspase 分子，及一系列诱发细胞发生凋亡的底物蛋白。基因剔除技术在鼠和果蝇中的体内研究试验表明：Apaf-1 在多种信号刺激诱发凋亡的过程中起着关键的作用[43]。

2．接头分子的种类（表 10-6）

表 10-6

Unique family of caspase activating molecules	种类	结构	作用
CED4	线虫（C. elegans）nematode	N-terminal CARD（caspase recruitment domain）followed by NB domain（nucleotide-binding）	与 CED3 酶原的 N 端 CARD 互补区结合，将前 CED3 水解成为大小亚基的异源四聚体（自身具有活性），以 ATP 依赖方式进行自身寡聚化，含有经典的 A、B 盒基序，在结合核苷三磷酸中发挥重要作用
Apaf-1	Human and other mammals	NB-ARC domain family（Apaf-1/R gene/CED4）或 NACHT domain，	与 CED4 的区别是：Apaf-1 的 NB-ARC 含有多 WD 重复区，参与 Apaf-1 的自我抑制，使之处于一种非活性的，非寡聚体化状态，直至与 Cyt-c 结合
R genes	Plants		A family of ATPase 与抗病相关

3. Apoptosome 中的重要组成分子（表 10-7）

表 10-7

species	constitution/subunits	function
Metazoan	three core components	
cells	interminglc in a ternary complex	
	CED-3	CED-3 kill cells
	CED-4	CED-4
	CED-9	CED-9 protect cells from death
Mammalian	EGL-1	EGL-1 induce cell death by
Cells	caspase 9	disrupting apoptosome
	Apaf-1	interaction and represent a
	Bcl-2/xl	primordial signal integrator
	(pre-apoptotic members)	for the apoptsome

（二）实验性研究

王晓东实验室多年以来在凋亡复合体的研究中一直处于领先地位，他们首先发现了 cytochrome c 在线粒体凋亡途径中的始动作用，并首次克隆到 Apaf-1 全长基因，这就为他们进一步对凋亡复合体的组成、功能研究奠定了坚实的基础。该室已有多篇相关文献发表于 Cell、JBC、JCB 等国际核心刊物上。

1. 体外实验——利用纯化的重组蛋白进行体外重建

(1) 表达系统构建

采用杆状病毒表达系统（baculovirus expression vector）构建 Apaf-1 选择性拼接表达系统，在昆虫体系 Sf21 中表达经过修饰的重组 Apaf-1 蛋白（通常带有 9-His 标签），可经 Mono Q 纯化柱和 Nickel 亲和柱或传统的色谱柱进行纯化。同理还可以得到 His-pre-caspase 9 纯化蛋白。将二者共孵育后获得 35kDa 和 10kDa 两种活性形式；随即与经过亲和纯化的、^{35}S 标记的前 caspase 3 进行孵育后，只有在 Apaf-1、前 caspase 9、dATP 和 cytochrome c 全部存在时，前 caspase 3 才能够被活化水解。本实验代表着凋亡复合体的活性水平。

王晓东小组采用上述表达系统获得羧基端带 His 标签的一组 Apaf-1 蛋白，通过 nickel 亲和柱纯化及 SDS-PAGE 分离后，检测出若干 Apaf-1 拼接体，并经过 RT-PCR 证实它们都在 HeLa 细胞中存在。在这些选择性拼接体中，Apaf-1L-

WD13 具有最稳定的和 dATP 依赖的 caspase 3 激活活性。

(2) dATP/ATP 结合与水解

纯化后的 Apaf-1 与 α-^{32}P 标记的 dATP/ATP 共孵育后，通过 nickel 亲和树脂和薄层色谱（TCL）分析后发现具有放射性活性的核苷酸的确与 Apaf-1 共沉淀下来，表明 Apaf-1 确实与核苷酸相结合。有趣的是，与之结合的核苷酸大多为 ADP或 dADP 而不是 dATP/ATP。这说明结合的核苷酸被水解了。与 cytochrome c 共孵育并不能影响核苷酸的结合与水解。采用非水解形式的 ATP 类似物 ATPγS 与 dATP/ATP、Apaf-1、前 caspase 9 共孵育，发现其以一种剂量依赖的方式抑制着 caspase3 的活化。从而证明与 dATP/ATP 结合并将其水解为 dADP/ADP，对 Apaf-1 发挥功能具有重要意义。

(3) caspase 9、Apaf-1 和 cytochrome c 的相互作用

将纯化后的 Apaf-1（带 His 标签）、前 caspase 9（不带标签）和 cytochrome c 以不同的组合共孵育，用这三种蛋白的抗体分别检测沉淀与上清。上清中未检出 Apaf-1 和 caspase 9；只有在 Apaf-1、dATP、cytochrome c 同时具备时才会在沉淀中将它们检出。有意思的是，绝大多数 caspase9 沉淀中已有 35kDa 活性形式存在，表明前 caspase 9 一旦与 Apaf-1 结合就会迅速地自我激活。当反应体系中缺少 Apaf-1、dATP 或 cytochrome c 时，前 caspase9 仍然存在于上清中。值得注意的是，一些经过加工的 caspase9 仍然保留在上清中，提示前体 caspase9 一经加工便立即被 Apaf-1 释放，而且无论 dATP 存在与缺乏，cytochrome c 都会与 Apaf-1 发生共沉淀，但 cytochrome c 的沉淀量与 dATP 的增长相一致，说明 dATP 可以稳定 cytochrome c 与 Apaf-1 的结合。

(4) 凋亡复合体三元复合物（caspase 9·Apaf-1·cytochrome c）的形成

采用 nickel 树脂或 caspase 9 抗体共沉淀的方法发现 caspase 9 在 cytochrome c 和 dATP 存在下与 Apaf-1 相互作用。为进一步了解该三元复合物的生化特性，在 dATP 存在或缺乏时，Xuejun 等分别以凝胶过滤和分子筛（sizing column）分离复合物中的组分并经对比后发现：无 dATP 时，三者分别在 14、16、21 组分处出现，它们的泳动速率与各自单体的分子质量标准相吻合；若预先以 dATP 孵育，则三种蛋白只在组分 11 处出现，表明它们已形成了大的复合物，此时，如果对前 caspase 9 进行加工，将在柱上出现两个峰，第一个峰在组分 11 的位置上出现，与复合物形式一致，第二个峰位于组分 16 位置，是加工后的 caspase 9 从复合物上被释放出来[44~46]。

2. 体内实验

(1) 复合体的纯化

为了解上述体外重组的大复合体（经凝胶过滤色谱柱纯化后相对分子质量＞1.3mDa）在体内凋亡发生时的形成情况，取正常生长的 HeLa 细胞和以 stau-

rosporine 诱导凋亡的细胞提取物，通过相同的凝胶过滤柱后，以 Apaf-1 的抗体进行 SDS-PAGE、Western blotting 分析，在正常细胞提取物中 Apaf-1 在组分 14 峰处出现，与其非活性单体的位置相一致。在凋亡细胞中，Apaf-1 移向组分 11 峰，与在大复合物中显示的结果一致[44]。

(2) Apaf-1 寡聚化分析

Kelvin Cai 等对 ATP 激活的细胞提取成分进行分析发现：通过分子截留色谱柱，即快速蛋白液相色谱蛋白纯化系统分析（16/60）、制备（26/60）High Prep S300 Sephacryl 高度溶解和 Superose 6 HR 10/30 柱（Amersham Pharmacia）分离得到两种含 Apaf-1 的复合物形式，无生物活性的 1.4MDa 和有生物活性的 700kDa 的凋亡复合体成分。经过时间依赖的形成与活性的动力学分析发现，700kDa 复合物在 dATP 激活后 5min 即达高峰；而 1.4MDa 复合物要在 20min 后才达到高峰。应用荧光分光光度计分析其 DEVDase 活性（即 caspase3、7 前体），在 dATP 激活后 5min700kDa 复合物就可以表现出较高的加工前体 caspase 从而产生效应 caspase 的催化活性。与之对应的是，1.4MDa 复合物在所有测定时间点上都没有表现出活性。而且在凋亡细胞中，700kDa 复合物占据优势。从而推断这种 700kDa 的 Apaf-1 寡聚化复合物是凋亡复合体的功能活性形式[32]。

(三) 应用性研究

1. 应激研究——分子伴侣与凋亡复合体

细胞应激反应可能是通过表达热激蛋白介导细胞保护效应，干扰细胞凋亡的发生。由应激所诱导的凋亡是通过线粒体释放 cytochrome c，与 Apaf-1 及 caspase 形成复合体，引发 caspase 级联效应而实现的。因而，分子伴侣（Hsp 家族）对于凋亡的作用日益受到重视。

(1) Hsp90 与凋亡

Pandey 等的实验研究表明：Hsp90 与 Apaf-1 形成胞质复合物，并由此抑制了活性凋亡复合体的形成，Hsp90 的免疫耗竭造成了 Apaf-1 的丢失；阻断了前 caspase 9 的活化。将纯化后的 Apaf-1 加入 Hsp90 免疫耗竭后的细胞质提取物中重新恢复了 cytochrome c 介导的前 caspase 9 的活化。另外，用多种 DNA 致损伤剂处理细胞可以解离 Hsp90 与 Apaf-1 形成的胞质复合物，进而消除其对于前 caspase 9 活化的抑制。上述结果是分子伴侣参加 caspase 级联的第一个证据，并首次阐明了 cytochrome c 依赖的细胞凋亡途径存在着负调节因子[40]。

(2) Hsp 70 与凋亡

Saleh 等发现，Hsp70 通过与 Apaf-1 的 CARD 结构域直接作用来阻止凋亡复合体的形成（具体说来是阻止其寡聚化的形成）及与前 caspase 的结合。由此，他们提出：应激细胞和某些肿瘤细胞可能是通过组成型高表达 Hsp70 调节 Apaf-

1 的功能来抑制凋亡的发生[42]。Beere 等却观察到 Hsp70 并不阻止 Apaf-1 的寡聚化、也不结合于 caspase 的前体，而是与 Apaf-1 发生作用，阻止其将 caspase 招募到凋亡复合体，进而阻断了功能复合体的组装[47]。

(3) Hsp27 与凋亡

Concannon 等发现，Hsp27 可以抑制 cytochrome c 介导的 caspase 3 的活化。从 cytochrome c 激活的细胞质中免疫耗竭掉 Hsp27 后，导致 caspase 活性的降低。免疫沉淀 Hsp27 产生了 cytochrome c 和前 caspase 3 的共沉淀；在交互实验中，cytochrome c 和前 caspase 3 的免疫沉淀也可导致 Hsp27 的共沉淀，表明这是两种各自独立的相互作用。上述结果显示：Hsp27 通过捕获并截留 cytochrome c 和前 caspase 3 来阻止 caspase 3 的活化，因此，妨碍了凋亡复合体的正确组装以及功能的发挥[48]。

2．肿瘤抗凋亡机制研究

蛋白酶解所依赖的细胞凋亡对于维持机体的稳态具有重要意义，一旦此通路失活将导致肿瘤的发生。因此，人们对凋亡复合体在肿瘤发生、发展过程中的作用进行了探讨。已有研究表明 Apaf-1 的活性与肿瘤凋亡密切相关。Wolf 等采用无细胞体系比较了正常卵巢上皮细胞（NOE）提取物与 10 种卵巢癌细胞系的细胞质提取物，发现几乎所有卵巢癌细胞系的 cytochrome c 依赖的蛋白酶活性都消失了。抗凋亡细胞系表达 Apaf-1 和 caspase 9、3、7；然而与 NOE 系相比均缺乏 Apaf-1 的活性。Apaf-1 的竞争性抑制剂可能与 Apaf-1 活性缺失有关，但他们并没有检测出明显的 Apaf-1 抑制剂的存在或 Apaf-1 异构体的表达、Apaf-1 缺失、降解及突变。这些研究提示调节 Apaf-1 的活性是卵巢肿瘤抗凋亡的一个重要机制[37]。紫外线诱导的肿瘤凋亡发生于凋亡途径中 caspase3 激活的上游、线粒体 cytochrome c 释放的下游。Jia 等发现人类白血病细胞系 K562、CEM、CEM/VLB（100）、人白血病母细胞和膀胱癌细胞系 J82 对紫外线诱导的凋亡具有不同的敏感性，而对凋亡的敏感性与 caspase 9b 或 XIAP（凋亡抑制蛋白）的表达或前 caspase 9 的水平无关。这一途径的阻断可能是 Apaf-1 缺乏的结果。转染 Apaf-1 至抗凋亡细胞中可以增加其敏感性。因此，Apaf-1 的缺乏是导致白血病抗凋亡的一个重要原因[36]。此外，凋亡通路中的重要“效应分子”—— caspase3 是肿瘤抗凋亡作用的重要环节。Friedrich 等将前 caspase3 和 caspase3 分别转染到具有获得性抗药性（MT1/ADR）的乳腺癌细胞 MCF-7 中，使之过度表达 caspase3，可以增强其对化疗的敏感性[49]。

3．Bcl-2 家族对凋亡复合体的调节

Bcl-2 蛋白家族是凋亡途径中非常关键的上游调节分子，分为抗凋亡（anti-apoptotic）和前凋亡（pro-apoptotic）两大类成员。抗凋亡成员如 Bcl-2、Bcl-xl 通

过抑制凋亡复合体的形成或阻止 cytochrome c 及 AIF 释放至胞浆来发挥作用。与之相反的是，前凋亡成员如 Bax、Bak 以异源二聚体的形式促进 caspase 从死亡拮抗剂中释放出来，并通过改变线粒体膜的通透性而有利于凋亡因子的释放，进而引发 caspase 的活化。因此，Bcl-2 家族蛋白在通用凋亡通路中是决定细胞生死存亡的关键所在[50]。Apaf-1 是这些 pro-survival 成员的重要靶分子，缺乏 Apaf-1 的细胞可以抵抗多种被 Bcl-2 所抑制的细胞毒性信号。然而，Bcl-2 与其家族成员是否直接控制着 Apaf-1 的活性始终是悬而未决的中心问题。过去曾认为：在死亡信号刺激的作用下，Bcl-2 和 Bcl-xl 通过抑制 Apaf-1 和共激活子 cytochrome c 由线粒体释放而抑制大的凋亡复合体的形成。Hausmann 通过对 Apaf-1 和 Bcl-2、Bcl-xl 进行亚细胞定位进而了解它们之间的相互作用。采用自制的 Apaf-1 单克隆抗体鉴定了从多种细胞（HeLa、MCF-7、293T、COS-7、Jurkat 及活化的人原始 T 细胞）中提取的组分，并进行激光共聚焦扫描显微镜检测后发现，Bcl-2 和 Bcl-xl 在亚细胞器膜上表达显著，而内源性 Apaf-1 在胞浆中含量丰富。无论这两种 pro-survival 蛋白在过表达或诱导凋亡发生后，二者与 Apaf-1 的共定位都不相一致。随后用免疫金标记的电镜检测以 UV 诱导的 HepG2 细胞凋亡模型，发现 Apaf-1 也是广泛分布于胞浆中而并非位于线粒体或其他细胞器内。另外，体外实验表明，Bcl-2 和 Bcl-xl 不能抑制外源 cytochrome c 对 Apaf-1 的激活。因此，根据上述结果可以初步否定 Bcl-2 样蛋白通过捕获并截留 Apaf-1 阻止凋亡发生的假说，并提出 Bcl-2 的抑制作用可能是通过控制活化上游的关键事件来对 Apaf-1 进行调节而实现[33]。

4．两个凋亡通路的衔接（crosstalk）

泛素-蛋白酶复合体对于 Bcl-2 的降解代谢十分关键，这一过程可以被蛋白酶抑制剂所阻断。同时模拟 Bcl-2 蛋白在 MAP 激酶位点（Thr56、Thr74、Ser87）的磷酸化也可防止它的降解。因此，Bcl-2 的脱磷酸化降解可以排除它对凋亡的抑制，并且对 caspase 级联的活化也具有放大作用。由此我们也可以看出 Bcl-2 作为蛋白酶复合体和凋亡复合体两大通路的重要交汇分子将二者有机地联系起来，使蛋白酶复合体不仅是凋亡复合体下游细胞蛋白溶解事件的主要执行者，而且还参与了对凋亡途径的负调控，从而使整个细胞的生命活动实现高度有序、协调和统一[51]。

最近的许多研究表明，死亡受体介导的凋亡通路成员 caspase 8 及其底物 Bid 在特定的信号刺激下，以一种非受体依赖的方式被激活，提示在这两大凋亡“信号臂”中存在着明显的 crosstalk，使它们彼此之间可以吸收对方的优势。Tang 等发现至少在一定的细胞类型和一定的条件下，caspase 8 的激活及随后 Bid 的裂解的确参与了 cytochrome c 介导的凋亡。而且活化 caspase 3 的参与对于 caspase 8 的激活和 Bid 的加工十分重要。尽管 caspase 8 活化并非 cytochrome c 介导的凋亡

途径所必需，但它能够显著缩短作用的时间。因此 caspase 8 参与到另一条凋亡途径有助于放大较弱的死亡信号，以保证凋亡事件能够在特定的时间框架下顺利执行[52]。

此外，在 TGF-β1 的作用下诱导的细胞凋亡也可以证实这两条途径存在着 crosstalk。TGF-β1 通过受体依赖的方式诱导肝细胞的凋亡、控制着肝脏的大小。Freathy 等在利用 TGF-β1 诱导鼠肝癌细胞凋亡时发现伴有细胞色素 c 的释放、Apaf-1 寡聚化形成 700kDa 的凋亡复合体以及 caspase2、3、7、8 的加工和活化。体外采用肝癌细胞提取物在 dATP 的参与下也形成了 700kDa 的凋亡复合体，并招募 caspase3 和 7。应用全蛋白酶抑制剂（Z-VAD. FMK）阻断了 dATP 刺激下的 caspase 活化，但并不影响 Apaf-1 凋亡复合体的组装。同样地，在完整细胞中，全蛋白酶抑制剂阻断了 TGF-β1 诱导的细胞凋亡以及 caspase3 和 7 的招募，但也不干扰 Apaf-1 凋亡复合体的组装。而且 Cain 也观察到类似的结果。从而证明了这两大途径之间确实存在密切的联系。对于众多不同的凋亡启动信号而言，线粒体/ Apaf-1/caspase 活化通路是共有的通用执行方式，在整个凋亡事件的调控中占据重要地位[39,51]。

五、信号复合体研究方法

（一）概述

COP9（constitutive photomorphogenic）signalosome（CSN），称作 COP9 信号复合体或含 JAB1（jun-activation-domain-binding-protein 1）的信号复合体，最初是在对拟南芥种子植入发育的光形态发生的遗传学分析中被发现，它是光形态发生的一个负调节子[53,54]。在哺乳动物和植物中，CSN 亚基的序列和亚基组成非常保守，表明它在高等真核生物中具有相当保守的细胞生物学功能[53-57]。美国耶鲁大学分子、细胞及发育生物学研究室华裔科学家邓兴旺教授所领导的实验室在信号复合体的研究方面始终处于世界前列，至今已有多篇文章发表于 Cell 等生命科学核心期刊[53-60]，许多新发现和关键实验方法的建立都来源于这个群体，涌现出这一领域相当一部分的研究精英。该实验室研究人员魏宁于 1994 年首次于 Cell 上报道“COP9 信号复合体是一个介导拟南芥光控发育的新的信号复合物”，以这种模式生物为对象由此拉开了该重要复合体研究的序幕[53]。德国学者 Seeger 于 1998 年采用网织红细胞裂解系统首次鉴定出这个含有 JAB1、相对分子质量大约为 450kDa 的新蛋白复合物，并基于它的激酶活性，将这个在人类中发现的与拟南芥 COP9 信号复合体同源的复合物命名为“含 JAB1 信号复合体”[55]。此外，以色列研究人员 Chamovitz 在特拉维夫大学建立了植物科学研究室，研究 COP9 信号复合体介导光信号转导调控植物发育的机制[54,59]。近年刚刚成立的由北京大学和耶鲁大学联合建立的“北京-耶鲁植物遗传和农业技术联

合中心”将我国这一领域的研究与世界接轨[61]。

信号复合体由 8 个不同的亚基组成，依据哺乳动物复合物亚基的相对分子质量，记为 S1（57kDa）－S8（22kDa）[53~57]（表 10-8）。1995 年在 Cell 第 82 期中报道了人类与拟南芥 COP9 复合体同源的亚基成分为 1（COP11、FUS6）、8（COP9）[62]。然而，有关该复合物是否在动植物中存在同源物的论断始终含混不清，直至近年来由两个不同的研究小组分别实现了从哺乳动物细胞中分离鉴定出 CSN 才使这一论断得以确证（表 10-9）。Seeger 等是从人红细胞中制备 26S 蛋白酶复合体时，作为污染成分意外获得了 CSN[63]，并随即将这两个复合体成功地共分离出来[55,64]。魏宁等在共富集亚基 1 和 8 的基础上[56]，采用生化方法从猪的脾脏中确定出该复合体，并利用抗亚基 1 的抗体，通过免疫亲合纯化法进一步证实了它的组成[57]。

表 10-8　信号复合体亚基组成

	Mammals		Arabidopsis			
Subunit	names	MW (kDa)	names	MW (kDa)	Homology (identity%)	Functional remarks
S1	GPS1	57	FUS6、COP11	51	42	Repressor of JNK, AP 1
S2	Trip15	51	—	54	60	*Nuclear receptor interacting Protein*
S3	SGN3	46	—	46	46	—
S4	—	45	—	47	60	—
S5	JAB1	38	AJH1	42	62	AP-1 coactivator
			AJH2	42	63	
S6	hVIP	36	—	36	50	cell-cycle regulator
S7a	—	30	—	—	—	—
S7b	—	30	FUS5	27	32	—
S8	hCOP9	23	COP9	22	32	—

表 10-9　信号复合体亚基命名

New name	Original nomenclature				
	Arabidopsis	Human	Mouse	Drosophila	S. pomb
CSN1	COP11, FUS6	Sgn1, GPS1	COPS1, Mfh	DCH1	Caa1, Sgn
CSN2	Subunit2	Sgn2, TRIP15 Halien	COPS2	DCH2, Alien	Sgn
CSN3	Subunit3	Sgn3	COPS3	DCH3	?
CSN4	COP8, FUS4	Sgn4	COPS4	DCH4	Sgn
CSN5	AJH1, AJH2	Sgn5, JAB1	COPS5, Jab1	DCH5	?
CSN6	Subunit6	Sgn6, hVIP	COPS6	DCH6	?
CSN7	FUS5	Sgn7	COPS7a, COPS7b	DCH7	?
CSN8	COP9	Sgn8, hCOP9	COPS8	DCH8	?

（二）研究方法

1．酵母双杂交分析

酵母双杂交是研究蛋白之间相互作用的经典方法，被广泛应用于各类筛选实验中，在进行复合体组成蛋白与结合蛋白相互作用分析时，是重要且有效的手段。Li 等在研究黄体生成素/促性腺激素的代谢时，利用鼠黄体生成素/促性腺激素受体（rLHR）C 末端尾巴为诱饵，经酵母双杂交鉴定出 p38JAB1 作为选择性 rLHR 结合物。以此为前提，经过一系列后续实验证明 p38JAB1 与 rLHR 前体相互作用，促进它的降解[65]。Chauchereau 等利用酵母与哺乳动物双杂交体系，并结合 pull-down 实验，确定 JAB1 与前列腺素受体（PR）和固醇受体共激活物 1（SRC-1）发生作用，并稳定了 PR-SRC-1 复合物[66]。Chamovitz 实验室也利用双杂交实验发现果蝇 COP9 复合体中蛋白成分之间部分通过 PCI 结构域发生作用。通过 P 元件插入（P-element insertion）或基因缺失法破坏其中一个亚基成分，导致在幼虫晚期或蛹期发生致死性改变。这种致死改变可能是各种基因多效性的影响所造成，从而反映了 COP9 复合体在动物发育过程中的重要作用[67]。

干扰素一致序列结合蛋白（ICSBP）是干扰素调节因子（IRF）家族的成员，在介导小鼠抗病原感染以及促进髓系分化中具有相当重要的作用。ICSBP 插入通过与其他因子作用而发挥其某些转录活性，使之能够结合于多种含有 DNA 合成成分的启动子元件。这种特定的结合通常是通过特定的 COOH 末端称作 IAD（IRF 相关结构域）介导。为拓展对于 ICSBP 与其他因子相互作用能力的了解，Cohen 等采用 ICSBP-IAD 为诱饵筛选 B 细胞 cDNA 文库，鉴定出 Trip15（新近被确定为 CSN 组分，CSN2）是在酵母细胞中与 ICSBP 发生特定相互作用的因子。体外 GST pull-down 实验和在 COS7 细胞中的免疫共沉淀实验均确证了这种结合的存在。进一步研究 CSN 与 ICSBP 相关性的体外实验表明，ICSBP 的 IAD 内部一个独特的丝氨酸残基被 CSN 磷酸化，而该磷酸化残基对于 ICSBP 通过 IRF-1 产生抑制活性十分关键。这一实验表明 CSN 在整合导入信号、影响 ICSBP 的转录活性方面具有重要作用[68]。

2．免疫共沉淀

Chamovitz 实验室从果蝇中分离出编码黑腹果蝇 COP9 复合体的 8 个亚基，通过免疫共沉淀并结合凝胶过滤分析确证了这些蛋白为 COP9 复合体的组分。为进一步研究它们之间的相互作用奠定了基础[67]。

3．亚细胞定位研究

这是确定蛋白间相互作用的最可靠方法之一，只有它们处于相同的亚细胞区

(compartment)，才有可能实现功能上的彼此相关。Seeger 在 1998 年研究 CSN 与 26S 蛋白酶体的盖子之间的相互关系时，利用免疫荧光染色（immunofuorescence staining）反映出这一新的蛋白复合体与 26S 蛋白酶体具有相似的亚细胞分布，由此进一步探讨了二者在调节转录活性功能上的相关性[55]。信号复合体在所有植物和动物所有器官中都可以检测到，似乎在拟南芥的花和根部以及鼠的胸腺、脾和脑表达水平稍高[54,69]。CSN 的亚细胞分布已被许多研究所报道。在植物和人类细胞中，采用 S1 或 S8 亚基的抗体检测出明显的细胞核定位[54,56,69]。人类的 S6（VIP）也显示为一个核蛋白[70]。另一方面，Sgn3（S3）大多位于胞质中，特别是围绕在核的周围[55]。然而，只有 S1 和 S8 被报道仅仅以复合体的形式聚集。因此它们的定位形式更能反映 CSN 的状况。与之形成对照的是，拟南芥 S5 和 S7 及人 S5 可以以复合体的形式以及游离单体的形式存在。由于拟南芥 S5 的单体形式大多位于胞质[71]，那么对细胞内一个给定的亚单位进行细胞定位可能并不能代表着整个复合体的情况。

4. 信号复合体的分离与鉴定[56]

（1）细胞粗提物

当细胞生长饱和后，以 PBS 洗涤细胞 3 次，刮除细胞，离心，以低渗缓冲液重悬细胞（10mmol/L HEPES、pH 7.6，1.5mmol/L、$MgCl_2$、10mmol/L KCl，含 0.2% NP40、5mmol/L DTT、2mmol/L PMSF），液氮反复冻融 3 次，以 4℃，12 000×g 离心 10min，上清用 0.2μm 滤膜过滤，然后通过 superdex 200 gel filtration 过滤柱（Amersham Pharmacia）进行复合体的纯化。

（2）从组织中提取

以小鼠肝脏为例，先于液氮中冷冻，用上述低渗溶液进行匀浆，样品离心 2 次，除去碎屑，上柱前过滤除渣，以凝胶过滤柱（superdex 200 gel filtration column）纯化复合体，收集洗脱液，1ml/份，以 Microprep 10 concentration（Filtron）或 Strataclean beads（Stratagen）进行浓缩。具体方法是：每 0.5ml 组分加入 25μl Strataclean beads，与样品混匀 10min，与非特异蛋白结合的珠子沉淀下来，收集上清，加入等体积的 SDS 样品缓冲液，最后进行 SDS-PAGE 以及标准蛋白印迹分析。

（3）信号复合体的大量生化纯化

以猪脾脏为例，将 190g 样品加入 250ml 提取缓冲液（50 mmol/L Tris-HCl，pH 7.0，1.5mmol/L $MgCl_2$，10mmol/L KCl，0.2mmol/L EDTA，5% 甘油，5mmol/L DTT，2mmol/L PMSF）以及蛋白酶抑制剂 cocktail（Rocher）进行匀浆，匀浆液经过 4 层纱布过滤，12 000g 离心 2 次，每次 15min，上清与聚乙二醇缓慢混合，直至终浓度为 12%，混匀后，4℃ 放置 20min，10 000g 离心 10min。沉淀以 120ml 缓冲液 B（50 mmol/L Bis-Tris pH 6.4，1.5mmol/L $MgCl_2$，

10mmol/L KCl，0.2mmol/L EDTA，0.01% NP40，100mmol/L NaCl，10% 甘油，5mmol/L DTT，2mmol/L PMSF）以及蛋白酶抑制剂 cocktail，4℃混合 2h，溶液于 12 000g 离心 15min，15 000g 离心 30min，上清上样于 300ml Q Sepharose fast flow column（Amersham Pharmacia），含有复合物的组分以 0.2～0.4mol/L NaCl 梯度的缓冲液 I（20 mmol/L Bis-Tris pH 6.4，10% 甘油）进行洗脱。随后，以 0.3mol/L NaCl、0.15mol/L Na_2SO_4 进行调整，上样于 60ml Phenyl Sepharose high performance column（Amersham Pharmacia），以缓冲液 II（10 mmol/L Bis-Tris pH7.0，10% 甘油）进行洗脱。样品调整至含有 0.1mol/L NaCl 并加样于 10ml 肝素亲和柱（Amersham Pharmacia），以 0.1mol/L 磷酸钠（pH7.0）、0.2mol/L NaCl 洗脱，同样以缓冲液 II 冲洗柱子。随后，将洗脱液上样于 1ml Mono Q column（Amersham Pharmacia），以 0.2～0.4mol/L NaCl 梯度的缓冲液 II 洗脱，收集含复合物组分，用 Microprep 10 concentration（Filtron）进行浓缩。最后，将 0.2ml 样品上样于 Superose 6 gel filtration column（Amersham Pharmacia），复合物洗脱速度为 0.3ml/min，收集组分，每份 0.5ml，以 Microprep 10 concentration 再次进行浓缩。上述所有步骤都应在自动化 HPLC 系统监控下，在 4℃实验室中进行。

参考文献

1 Grabowski P J, Seiler S R, Sharp P A. A multicomponent complex is involved in the splicing of messenger RNA precursors. Cell, 1985, 42 (1): 345～353

2 Brody E, Abelson J. The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. Science, 1985, 228 (4072): 963～967

3 Ajuh P, Kuster B, Panov K, Zomerdijk J C, Mann M, Lamond A I. Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. EMBO J, 2000, 19 (23): 6569～6581

4 Awasthi S, Palmer R, Castro M, Mobarak C D, Ruby S W. New roles for the Snp1 and Exo84 proteins in yeast pre-mRNA splicing. J Biol Chem, 2001, 25 [epub ahead of print]

5 Ambrozkova M, Puta F, Fukova I, Skruzny M, Brabek J, Folk P. The fission yeast orthologue of the coregulator SKIP interacts with the small subunit of U2AF. Biochem Biophys Res Commun, 2001, 284 (5): 1148～1154

6 Makarova O V, Makarov E M, Luhrmann R. The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. EMBO J, 2001, 20 (10): 2553～2563

7 Blencowe B J, Bauren G, Eldridge A G, Issner R, Nickerson J A, Rosonina E, Sharp P A. The SRm160/300 splicing coactivator subunits. RNA, 2000, 6 (1): 111～120

8 Labourier E, Rio D C. Purification of Drosophila snRNPs and characterization of two populations of functional U1 particles. RNA, 2001, 7 (3): 457～470

9 Nagai K, Muto Y, Pomeranz Krummel D A, Kambach C, Ignjatovic T, Walke S, Kuglstatter A. Structure and Assembly of the Spliceosomal snRNPs. Biochem Soc Trans, 2001, 29 (2): 15～26

10 Newby M I, Greenbaum N L. A conserved pseudouridine modification in eukaryotic U2 snRNA induces a

change in branch-site architecture. RNA, 2001, 7 (6): 833～845

11 Hershko A, Ciechanover A. The ubiquitin system. Annu Rev Biochem, 1998, 67: 425～479

12 Laney J D, Hochstrasser M. Substrate targeting in the ubiquitin system. Cell, 1999, 97: 427～430

13 Schwartz A L, Ciechanover A. The ubiquitin-proteasome pathway and pathogenesis of human diseases. Annu Rev Med, 1999, 50: 57～74

14 Baumeister W, Walz J, Zuhl F, Seemuller E. The Proteasome: Paradigm of a Self-Compartmentalizing Protease. Cell, 1998, 92: 367～380

15 Kisselev A F, Akopian T N, Castillo V, Goldberg A L. Proteasome active sites allosterically regulate each other, suggesting a cyclical bite-chew mechanism for protein breakdown. Mol Cell, 1999, 4: 395～402

16 Voges D, Zwickl P, Baumeister W. The 26S Proteasome: a molecular machine designed for controlled proteolysis. Annu Rev Biochem, 1999, 68: 1015～1068

17 Verma R, Chen S, Feldman R, Schieltz D, Yates J, Dohmen J, Deshaies R J. Proteasomal Proteomics: Identification of Nucleotide-sensitive Proteasome-interacting Proteins by Mass Spectrometric Analysis of Affinity-Purified Proteasomes. Mol Bio Cell, 2000, 11: 3394～3425

18 Low P, Hastings R A, Dawson S P, Sass M, Billett M A, Mayer R J, Reynolds S E. Localization of 26S proteasomes with different subunit composition in insect muscles undergoing programmed cell death. Cell Death Differ, 2000, 7 (12): 1210～1217

19 Mykles D L. Purification and characterization of a multicatalytic proteinase from crustacean muscle: comparison of latent and heat-activated forms. Arch Biochem Biophys, 1989, 274 (1): 216～228

20 Farout L, Lamare M C, Cardozo C, Harrisson M, Briand Y, Briand M. Distribution of proteasomes and of the five proteolytic activities in rat tissues. Arch Biochem Biophys, 2000, 374: 2207～2212

21 Khan M, Joseph S. Characterization of Membrane-Associated Proteasomes in WB Rat Liver Epithelial Cells. Arch Biochem Biophys, 2001, 385 (1): 99～107

22 Briane D, Olink-Coux M, et al. Immunolocalization of a specific type of proteasome close to the bile canaliculi in fetal and adult rat liver. Eur J Cell Biol, 1992, 57 (1): 30～39

23 Briane D, Olink-Coux M, Vassy J, Oudar O, Huesca M, Scherrer K, Fourcrier J. Immunolocalization of a specific type of proteasome close to the bile canaliculi in fetal and adult rat liver. Eur J Cell Biol, 1992, 57 (1): 30～39

24 Piccinini M, Tazartes O, Mostert M, Musso A, DeMarchi M, Rinaudo M T. Structural and functional characterization of 20S and 26S proteasomes from bovine brain. Brain Res Mol Brain Res, 2000, 76: 1103～1114

25 Nelson J E, Loukissa A, Altschuller F C, Monaco J J, Fallon J T, Cardozo C. Up-regulation of the proteasome subunit LMP7 in tissues of endotoxemic rats. J Lab Clin Med, 2000, 135: 4324～4331

26 Kabore A F, Wang W J, Russo S J, Beers M F. Biosynthesis of surfactant protein C: characterization of aggresome formation by EGFP chimeras containing propeptide mutants lacking conserved cysteine residues. J Cell Sci, J, 2001, 114 (Pt 2): 293～302

27 Kopito R R. Aggresomes, inclusion bodies and protein aggregation. Trends Cell Biol, 2000, 10 (12): 524 ·530

28 Johnston J A, Ward C L, Kopito R R. Aggresomes: a cellular response to misfolded proteins. J Cell Biol, 1998, 143 (7): 1883 ·1898

29 Waelter S, Boeddrich A, Lurz R, Scherzinger E, Lueder G, Lehrach H, Wanker E E. Accumulation of mutant Huntingtin fragments in aggresome-like inclusion bodies as a result of insufficient protein degradation. Mol Biol Cell, 2001, 12 (5): 1393～1407

30 Dul J L, Davis D P, Williamson E K, Stevens F J, Argon Y. Hsp70 and antifibrillogenic peptides promote degradation and inhibit intracellular aggregation of amyloidogenic light chains. J Cell Biol, 2001, 152 (4): 705～716

31 Garcia-Mata R, Bebok Z, Sorscher E J, Sztul E S. Characterization and dynamics of aggresome formation by a cytosolic GFP-chimera. J Cell Biol, 1999, 146 (6): 1239～1254

32 Cain K, Bratton S B, Langlais C, Walker G, Brown D G, Sun X M, Cohen G M. Apaf-1 Oligomerizes into Biologically Active ～700-kDa and Inactive～1.4-MDa Apoptosome Complexes. J Biol Chem, 2000, 275 (9): 6067～6070

33 Hausmann G, O'Reilly L A, van Driel R, Beaumont J G, Strasser A, Adams J M, Huang D C. Pre-apoptotic Apoptosis Protease-activating Factor 1 (Apaf-1) Has a Cytoplasmic Localization Distinct from Bcl-2 or Bcl-xL. J Cell Biol, 2000, 149 (3): 623～633

34 Tang D, Lahti J M, Kidd V J. Caspase-8 activation and bid cleavage contribute to MCF7 cellular execution in a caspase-3-dependent manner during staurosporine-mediated apoptosis. J Biol Chem, 2000, 275 (13): 9303～9307

35 Lauber K, Appel H A, Schlosser S F, Gregor M, Schulze-Osthoff K, Wesselborg S. The adapter protein Apaf-1 is proteolytically processed during apoptosis. J Biol Chem, [epub ahead of print] 2001

36 Jia L, Srinivasula S M, Liu F T, Newland A C, Fernandes-Alnemri T, Alnemri E S, Kelsey S M. Apaf 1 protein deficiency confers resistance to cytochrome c-dependent apoptosis in human leukemic cells. Blood, 2001, 98 (2): 414～421

37 Wolf B B, Schuler M, Li W, Eggers-Sedlet B, Lee W, Tailor P, Fitzgerald P, Mills G B, Green D R. Defective cytochrome c-dependent caspase activation in ovarian cancer cell lines due to diminished or absent Apaf-1 activity. J Biol Chem, [epub ahead of print] 2001

38 Purring-Koch C, McLendon G. Cytochrome c binding to Apaf-1: the effects of dATP and ionic strength. Proc Natl Acad Sci U S A, 2000, 97 (22): 11928～11931

39 Freathy C, Brown D G, Roberts R A, Cain K. Transforming growth factor-beta (1) induces apoptosis in rat FaO hepatoma cells via cytochrome c release and oligomerization of Apaf-1 to form an approximately 700-kd apoptosome caspase-processing complex. Hepatology, 2000, 4 (1): 750～760

40 Pandey P, Saleh A, Nakazawa A, Kumar S, Srinivasula S M, Kumar V, Weichselbaum R, Nalin C, Alnemri E S, Kufe D, Kharbanda S. Negative regulation of cytochrome c-mediated oligomerization of Apaf-1 and activation of procaspase-9 by heat shock protein 90. EMBO J, 2000, 19 (16): 4310～4322

41 Jiang X, Wang X. Cytochrome c promotes caspase-9 activation by inducing nucleotide binding to Apaf-1. J Biol Chem, 2000, 275 (40): 31199～31203

42 Saleh A, Srinivasula S M, Balkir L, Robbins P D, Alnemri E S. Negative regulation of the Apaf-1 apoptosome by Hsp70. Nat Cell Biol, 2000, 2 (8): 476～483

43 Chu Z L, Pio F, Xie Z, Welsh K, Krajewska M, Krajewski S, Godzik A, Reed J C. A novel enhancer of the Apaf1 apoptosome involved in cytochrome c-dependent caspase activation and apoptosis. J Biol Chem, 2001, 276 (12): 9239～9245

44 Zou H, Li Y, Liu X, Wang X. An Apaf 1.cytochrome c multimeric complex is a functional apoptosome that activates procaspase-9. J Biol Chem, 1999, 274 (17): 11549～11556

45 Xue J, Wang X D. Cytochrome c Promotes Caspase-9 Activation by Inducing Nucleotide Binding to Apaf-1. J Biol Chem, 2000, 275 (40): 31199～31203

46 Genini D, Budihardjo I, Plunkett W, Wang X, Carrera C J, Cottam H B, Carson D A, Leoni L M. Nucleotide Requirements for the in Vitro Activation of the Apoptosis Protein-activating Factor-1-mediated Cas-

pase Pathway. J Biol Chem, 2000, 275 (1): 29～34

47 Beere H M, Wolf B B, Cain K, Mosser D D, Mahboubi A, Kuwana T, Tailor P, Morimoto R I, Cohen G M, Green D R. Heat-shock protein 70 inhibits apoptosis by preventing recruitment of procaspase-9 to the Apaf-1 apoptosome. Nat Cell Biol, 2000, 2 (8): 469～475

48 Concannon C G, Orrenius S, Samali A. Hsp27 inhibits cytochrome c-mediated caspase activation by sequestering both pro-caspase-3 and cytochrome c. Gene, 2001, 9 (4-5): 195～201

49 Friedrich K, Wieder T, Von Haefen C, Radetzki S, Janicke R, Schulze-Osthoff K, Dorken B, Daniel P T. Overexpression of caspase 3 restores sensitivity for drug-induced apoptosis in breast cancer cell lines with acquired drug resistance. Oncogene, 2001, 20 (22): 2749～2760

50 Tsujimoto Y. Role of Bcl-2 family proteins in apoptosis: apoptosomes or mitochondria? Genes Cells, 1998, 3 (11): 697～707

51 Dimmeler S, Breitschopf K, Haendeler J, Zeiher A M. Dephosphorylation targets Bcl-2 for ubiquitin-dependent degradation: a link between the apoptosome and the proteasome pathway. J Exp Med, 1999, 189 (11): 1815～1822

52 Tang D, Lahti J M, Kidd V J. Caspase-8 activation and bid cleavage contribute to MCF7 cellular execution in a caspase-3-dependent manner during staurosporine-mediated apoptosis. J Biol Chem, 2000, 275 (13): 9303～9307

53 Wei N, Chamovitz D A, Deng X W. Arabidopsis COP9 is a component of a novel signaling complex mediating light control of development. Cell, 1994, 78: 117～124

54 Chamovitz D A, Wei N, Osterlund M T, von Arnim A G, Staub J M, Matsui M, Deng X W. The COP9 complex, a novel multisubunit nuclear regulator involving in light control of a plant developmental switch. Cell, 1996, 86: 115～121

55 Seeger M, Kraft R, Ferrell K, Bech-Otschir D, Dumdey R, Schade R, Gordon C, Naumann M, Dubiel W. A novel protein complex involved in signal transduction possessing similarities to 26S proteasome subunits. FASEB J, 1998, 12 (6): 469～478

56 Wei N, Deng X W. Characterization and purification of the mammalian COP9 complex, a conserved nuclear regulator initially identified as a repressor of photomorphogenesis in higher plants. Photochem Photobiol, 1998, 68 (2): 237～241

57 Wei N, Tsuge T, Serino G, Dohmae N, Takio K, Matsui M, Deng X W. The COP9 complex is conserved between plants and mammals and is related to the 26S proteasome regulatory complex. Curr Biol, 1998, 8 (16):919～922

58 Wei N, Deng X W. COP9: a new genetic locus involved in light-regulated development and gene expression in arabidopsis. Plant Cell, 1992, 4 (12): 1507～1518

59 Wei N, Kwok S F, von Arnim A G, Lee A, McNellis T W, Piekos B, Deng X W. Arabidopsis COP8, COP10, and COP11 genes are involved in repression of photomorphogenic development in darkness. Plant Cell, 1994, 6 (5): 629～643

60 Wei N, Deng X W. The role of the COP/DET/FUS genes in light control of Arabidopsis seedling development. Plant Physiol, 1996, 112 (3): 871～878

61 Kang D, Wang X, Cao K, Sun C, Deng X W, Wei N. A gain-of-function phenotype conferred by over-expression of functional subunits of the COP9 signalosome in Arabidopsis. Plant J, 2000, 23 (5): 597～608

62 Seeger M, Ferrell K, Dubiel W. The 26S proteasome: a dynamic structure. Mol Biol Rep, 1997, 24 (1-2): 83～88

63 Spain B H, Bowdish K S, Pacal A R, Staub S F, Koo D, Chang C Y, Xie W, Colicelli J. Two human

cDNAs, including a homolog of Arabidopsis FUS6 (COP11), suppress G-protein and mitogen-activated protein kinase-mediated signal transduction in yeast and mammalian cells. Mol. Cell. Biol, 1996, 16: 6698－6706

64 Lee J W, Choi H S, Gyuris J, Brent R, Moore D D. Two classes of protein dependent on either the presence or absence of thyroid hormone for interaction with the thyroid hormone receptor. Mol Endocrinol, 1995, 9: 243－253

65 Li S, Liu X, Ascoli M. p38JAB1 binds to the intracellular precursor of the lutropin/choriogonadotropin receptor and promotes its degradation. J Biol Chem, 2000, 275 (18): 13386－13393

66 Chauchereau A, Georgiakaki M, Perrin-Wolff M, Milgrom E, Loosfelt H. JAB1 interacts with both the progesterone receptor and SRC-1. J Biol Chem, 2000, 275 (12): 8540－8548

67 Freilich S, Oron E, Kapp Y, Nevo-Caspi Y, Orgad S, Segal D, Chamovitz D A. The COP9 signalosome is essential for development of Drosophila melanogaster. Curr Biol, 1999, 9 (20): 1187－1190

68 Cohen H, Azriel A, Cohen T, Meraro D, Hashmueli S, Bech-Otschir D, Kraft R, Dubiel W, Levi B Z. Interaction between interferon consensus sequence-binding protein and COP9/signalosome subunit CSN2 (Trip15). A possible link between interferon regulatory factor signaling and the COP9/signalosome. J Biol Chem, 2000, 275 (50): 39081－39089

69 Derijard B, Hibi M, Wu I H, Barrett T, Su B, Deng T, Karin M, Davis R J. JNK1: a protein kinase stimulated by UV light and Ha-Ras that binds and phosphorylates the JUN activation domain. Cell, 1994, 76: 1025－1037

70 Glickman M H, Rubin D M, Coux O, Wefes I, Pfeifer G, Cjeka Z, Baumeister W, Fried V A, Finley D. A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9 signalosome and eIF3. Cell, 1998, 94: 615－623

71 Staub J M, Wei N, Deng X W. Evidence for FUS6 as a component of the nuclear-localized COP9 complex in Arabidopsis. Plant Cell, 1996, 8: 2047－2056

（姜 颖 王 研）

第十一章 蛋白质间连锁图的建立

第一节 酵母双杂交技术及其应用

一、酵母双杂交技术及其历史

酵母双杂交（yeast two-hybrid）技术是体内研究蛋白质相互作用的方法。它利用转录因子组件式（modular）结构的性质。这些因子常由两个以上相互独立的结构域组成，其中有DNA结合结构域（DNA binding domain，简称BD）和转录激活结构域（activation domain，简称AD），它们是转录因子发挥功能所必需的。分别与BD和AD融合的两个蛋白质（称为“诱饵”和“猎物”），如果在酵母的细胞核内发生相互作用，BD和AD相互靠近能恢复该转录因子的活性，激活报道基因的转录，从而使营养缺陷型的酵母细胞能在选择培养基上生长或具有其他可以检测的表型。1989年，Fields等人的工作标志着酵母双杂交系统的正式建立[1]。经过十余年的不断完善和发展，已经成为蛋白质组学中分析蛋白质间相互作用的不可或缺的工具。

最初的酵母双杂交系统是以转录因子Gal4为基础的，依靠招募RNA聚合酶Ⅱ激活转录，有些转录激活蛋白不能用此系统分析。因此，1997年由Marsoliter等人发展了以RNA聚合酶Ⅲ为基础的双杂交体系[2]，激活聚合酶Ⅱ的组件并不能激活聚合酶Ⅲ起始的转录，它有更低的假阳性率。在此系统中，聚合酶Ⅲ反应的报道基因SNR6的启动子被修饰，其中包含一个Gal4结合位点，有利于Gal4融合的诱饵结合于SNR6启动子。猎物用待分析的蛋白质与TFⅢC因子的τ138亚单位融合，τ138亚单位能招募聚合酶Ⅲ转录机器的其他成分。Gal4融合的诱饵和τ138亚单位融合的猎物蛋白质相互作用导致聚合酶Ⅲ反应的SNR6报道基因的转录。

在利用正向的双杂交筛选出大量的蛋白质相互作用的同时，蛋白质相互作用位点以及影响蛋白质结合、解离的因素成为新的研究热点。因此，人们发展了反向双杂交系统（reverse two-hybrid system），可以从大量的突变体中直接筛选出导致蛋白质间相互作用减弱或消失的突变位点。它的原理在于构建一种反向筛选的报道基因，蛋白质间相互作用激活报道基因表达，使细胞不能存活。因此，当两个相互作用的蛋白质由于某种原因不再相互作用时，细胞才能存活下来。常用的反向筛选的报道基因是URA3，它的基因产物为尿嘧啶合成所必需，但能将5-

氟乳清酸（5-FOA）转化为一种有毒的物质[3]，在添加了 5-氟乳清酸的培养基中，存在相互作用的细胞会因毒素的作用而死亡，只有已发生了蛋白质间解离或部分解离的转化株才可以生存。另一种反向筛选的体系称为分离杂交系统（split-hybrid system），利用了两个互相整合的报道基因，第一个报道基因含有大肠杆菌 tetR 抑制因子的编码序列，其表达受上游 lexA 操纵子调控；第二个报道基因是 His3，启动子区含有 tet 操纵子序列[4]。分别与 BD 和 AD 融合的两个蛋白质在酵母细胞核内如果发生相互作用，将激活 tetR 的表达，tetR 和 tet 操纵子序列的结合进一步影响 His3 基因的表达，只有不能发生相互作用的蛋白质才能用这个系统在组氨酸缺陷的培养基上筛选出来。除了筛选突变株获得蛋白质结合的信息外，反向双杂交更有意义的应用在于发现可导致相互作用的蛋白质间发生解离的药物或其他小分子物质。许多疾病与蛋白质异常的相互作用有关，发现具有促进解离的物质对于治疗疾病是非常有意义的。

上述的双杂交系统必须将相互作用的蛋白质共定位到细胞核内，有些蛋白质如膜相关蛋白不能采用这些系统分析。为了克服这些缺陷，有人发展了在细胞质中的双杂交体系，如 SOS 招募系统（SOS recruitment system），它灵活地运用了 Ras 信号转导途径的基本性质。酵母细胞需要 Ras 信号才能生存。Ras 的功能依赖于激活鸟苷酸交换因子（GEF），它把无活性 GDP-Ras 转变成激活的 GTP-Ras 形式。酵母中天然的 GEF 是 CDC25，它存在温度敏感的突变体。利用人的鸟苷酸交换因子 hSos 蛋白可以代替温度敏感型突变酵母的 RasGEF，即 CDC25，使酵母在 36℃可以存活。但是 hSos 蛋白必须定位到膜上才能刺激酵母 Ras 的鸟苷酸交换。实际应用中，一种蛋白质和 hSos 蛋白融合，另一种蛋白质和 Src 豆蔻酰化膜定位信号融合，如果两种蛋白质发生相互作用，就会将 hSos 招募到膜上，从而激活 Ras，使细胞在限制温度下存活和增殖[5]。虽然双杂交系统鉴定了一些膜外受体和配体的相互作用，但没有一个通用的方法适合所有的受体-配体相互作用，新的细胞质系统可能会更适合分析这样的相互作用。

为了克服酵母双杂交系统的某些缺陷，寻找更好的系统来研究蛋白之间的相互作用，人们也开始用蛋白质其他的结构特点来建立新的双杂交系统。分离的泛素系统（split-ubiquitin system）利用泛素的功能特点，当分别和泛素 N 端和 C 端部分融合的两个蛋白质相互作用时，使分离的泛素的两部分靠近，泛素专一性的蛋白酶就会识别泛素，导致报道蛋白的解离释放[6]。报道基因采用转录激活因子来激活报道基因的转录或其他易被检测到信号的蛋白质。同普通的利用转录因子结构特点的双杂交系统比较，这一系统有许多优点。第一，蛋白质相互作用水解释放报道蛋白，这样报道蛋白如果是转录因子，就更容易进入核内，活化转录。第二，报道蛋白可以选用其他的蛋白质如酶，这样可以通过分析酶的活性分析蛋白质间的相互作用。此外，还有一些利用酶的性质的双杂交系统，相互作用可以恢复酶的活性。

除研究蛋白质间相互作用的系统外，还发展了研究 DNA 和蛋白质间相互作用的系统——单杂交系统（one hybrid system）。单杂交系统用来鉴定与特定 DNA 序列直接作用的蛋白质[7]。“猎物”文库被用来检测与调控基序结合激活转录的蛋白质。还有研究 RNA 或小分子配体和蛋白质间的相互作用的系统——三杂交系统（three hybrid system）。三杂交系统引入了第三个相互作用的成分，如 RNA 或小分子配基作介导，三者共同作用，才能激活转录。SenGupta（1999）等[8]设计的酵母三杂交体系中把与 BD 融合的已知功能的 RNA 结合蛋白质——铁调节蛋白质 1（IRP1）作为“鱼钩”，将待研究的 RNA 片段与 IRP1 特异结合序列 IRE 连接成一个新的杂合序列作为“诱饵”，用 RNA 聚合酶Ⅲ的作用在酵母中表达；再将待筛选的蛋白质和 AD 连接，用与双杂交系统相似的方法，可以发现 RNA 特异结合的蛋白质或与蛋白质相互作用的 RNA。研究小分子配基与蛋白质间相互作用的三杂交系统与此类似，只是向酵母中引入小分子配基杂合体时，因为没有生物体系可以利用，需借助有机化学合成技术，而且亲疏水性的差异可能会使研究对象的范围受到一定的限制。这类三杂交系统有很多的用途，但也存在许多问题，如 RNA 片段的构象会因环境而改变，影响系统的检测，产生许多假阳性和假阴性。

在过去的几年中，酵母双杂交系统经过了许多改进，以便有更广泛的用途。例如，高等真核生物的许多蛋白质的翻译后修饰对于其功能是必需的，但在酵母中缺乏相关的修饰酶。Osborne（1995）等人[9]首先在系统中引入酪氨酸激酶，使其和诱饵共表达，然后筛选和磷酸化的诱饵专一相互作用的猎物文库。另外还有在双杂交系统中加入一种桥蛋白来搭桥、抑制或稳定蛋白质间的相互作用。而双诱饵的杂交系统允许同时研究一个蛋白质和两个诱饵之间的作用。

酵母双杂交系统的效用促进了哺乳动物细胞中检测蛋白质相互作用的类似系统的发展。这类系统有一些潜在的优势，如检测依赖翻译后修饰的相互作用。进一步说，可以检测和鉴定细胞刺激反应变化的相互作用。而在细菌中应用的双杂交系统可以使筛选更方便、快速和节省经费，因为分子生物学的技术更适合于细菌。细菌的转化率更高，可以覆盖更复杂的基因组文库和 cDNA 文库。在哺乳动物和细菌中应用的双杂交系统都已经商品化。

各种杂交系统都有自己的优缺点。但各种系统都不可避免的产生“假阳性”和“假阴性”。有些是系统本身造成的，如双杂交系统研究蛋白质复合体时，由于没有复合体中其他蛋白质的作用，多肽链会暴露出生理状态下为其他蛋白质覆盖的某些位点，或某一蛋白质不足以与作用蛋白形成稳定的结合，导致产生许多的假阳性和假阴性。值得注意的是，双杂交系统并不能解决所有的蛋白质相互作用问题。有些蛋白质不适合于这种方法。用 BD 和 AD 与目的蛋白融合可能改变了“诱饵”和“猎物”的构象，从而进一步改变了它们的功能。酵母双杂交系统一个主要的缺点是一些依靠翻译后修饰的相互作用在酵母中不能检测到。在真核

生物中这些修饰是常见的，如形成二硫键、糖基化和最普遍的磷酸化。虽然一些新的双杂交系统试着通过共表达负责翻译后修饰的酶来解决该问题，但还不能解决所有的翻译后修饰问题。因此双杂交系统需要进一步的改进，其他的研究蛋白质相互作用的手段也是必不可少的。

二、酵母双杂交技术在蛋白质组学中的应用

以上介绍了酵母双杂交的历史及主要的发展，在实际中酵母双杂交得到了很广泛的应用，尤其是在大规模分析蛋白质间相互作用的蛋白质组学的研究中的地位不可替代。全基因组序列的获得促进了大规模研究蛋白之间的相互作用，全面详尽的蛋白质相互作用图谱被称为互作组（interactome）。在病毒、细菌、酵母和线虫中都已经有大规模的酵母双杂交分析的报道。

T7 噬菌体被第一个用于大规模的双杂交分析。先构建随机的诱饵和猎物文库。每次用 10 个随机的非自激活的诱饵筛选随机的噬菌体蛋白质片段文库，接着用特定的诱饵筛选文库。在 55 个噬菌体蛋白中，共发现了 25 个相互作用，其中有 4 个已经有报道。这个研究表明，系统鉴定基因组编码特定蛋白质间众多的相互作用是可能的[10]。丙型肝炎病毒（HCV）能导致严重的肝病包括肝癌，但 HCV 的疫苗还没有开发出来。Flajolet（2000）等人研究了丙型肝炎病毒（HCV）中的蛋白质相互作用。丙型肝炎病毒的基因组编码一个多蛋白质，最终被加工成为 10 个成熟的蛋白质。用成熟的蛋白质分别构建“猎物”和“诱饵”表达载体，进行双杂交分析，但是没有检测到相互作用，可能是由于蛋白质折叠的问题。构建“诱饵”和“猎物”随机的 HCV 基因组文库。用随机筛选的 200 个诱饵和猎物文库双杂交分析，最后得到了 5 个相互作用的结果，其中 2 个已有报道。通过调节这些相互作用有可能开发出特异的抗病毒试剂[11]。用酵母双杂交分析痘苗病毒（vaccinia virus），先构建了痘苗病毒 266 个可读框的 BD 和 AD 阵列。用接合的方法筛选，在约 70 000 个组合中，共发现了 37 个相互作用，9 个是已知的。这些相互作用可以分为 DNA 复制、转录、病毒结构和形态发生、病毒和宿主相互作用和未知功能 5 个类别，提供了在特定的过程中未知蛋白质作用的线索[12]。

致人胃病的幽门螺旋菌（*Helicobacter pylori*）的序列在 1997 年已经测序完成，但对整个基因组序列编码蛋白质的功能知道的还很少。Rain（2001）等人[13]用酵母双杂交研究其蛋白质相互作用。它也是第一种用酵母双杂交技术大规模研究的原核生物。对 261 种 *H. pylori* 蛋白质（其中有 50 个和复合体、致病有关的蛋白质，211 个随机选择的蛋白质）和随机的 *H. pylori* 基因组文库进行杂交，得到了 1200 多个相互作用结果。其中有许多在大肠杆菌中也有类似的相互作用。筛选到的相互作用的蛋白质结构域将来可能有助于构建显性失活的突变体或用于新药设计。

酿酒酵母（*Saccharomyces cererisiae*）的全基因组序列测定在 1996 年完成，它共有 6000 多个可读框，用酵母双杂交大规模分析了这 6000 多个开放阅读框之间的相互作用。一种方法是阵列筛选法，192 个“诱饵”菌株和 6000 个“猎物”菌株一一结合，形成二倍体，结果筛选到了 281 个相互作用。另一种方法是文库筛选法，用约 6000 个“诱饵”菌株分别和含有 6000 种“猎物”的文库接合，共筛选到 692 个相互作用[14]。类似的工作中，Ito（2001）等人[15]分别克隆了酵母中几乎所有的可读框，分为 62 个“集合”（pools），每个“集合”含有近 96 个不同的克隆。62 个“诱饵”“集合”和 62 个“猎物”“集合”分别一一接合，得到了 1533 个相互作用的结果。其中有 841 个最可靠的相互作用，和前一结果相比，只有 141 个结果是共同的。这可能是由于 PCR 扩增中的突变消除了相互作用、质粒构建的特殊性影响蛋白质的折叠、或由于双杂交系统内在原因产生假的信号，也提示在像酵母这样较低等的真核生物中，蛋白之间的相互作用也是非常复杂的，用一种方法不可能筛选出所有的蛋白质相互作用。分析蛋白质相互作用的结果得到了几个有意义的网络，如与自体吞噬、纺锤极体的功能和囊泡运输有关的网络[15]。最近应用大规模的双杂交筛选技术，Drees（2001）等人[16]报道了用这种方法筛选与酵母中细胞极性有关蛋白的研究结果。用一个已知与细胞极性有关的 68 个蛋白作为诱饵，筛选一个阵列，这个阵列表达了约 90% 预测的酵母开放阅读框。他们共发现了 128 个新的蛋白质相互作用，其中 44 个包含未知功能的蛋白。他们最大的发现是在细胞信号通路之间新的联系，提示在已知的信号途径中存在联系，例如 Rho1 和 Cdc42 细胞途径，这些信号途径将肌动蛋白和形态建成检查点结合了起来。

在线虫（*Caenorhabditis elegans*）中研究了与 27 个生殖器发育有关的蛋白质相互作用的蛋白质，用 27 个诱饵筛选线虫的 cDNA 文库，共得到了 148 个相互作用结果，为深入研究线虫生殖器发育的分子机制提供了线索[17]。

三、蛋白质间连锁图的建立

分析蛋白质相互作用的目的是为了构建蛋白质间相互作用的连锁图。因为蛋白质相互作用控制着生命的各个过程。蛋白质间相互作用是细胞生命活动的基础和特征。例如：代谢和信号途径、形态发生途径、结构复合体和分子机器、DNA 的合成、基因转录激活、蛋白质翻译、修饰和定位、信号转导、细胞周期调控、中间代谢等重要的生物过程都涉及到蛋白质的相互作用。蛋白之间的相互作用、相互协调是细胞进行信号转导及一切代谢活动的基础，对蛋白质相互作用连锁关系的揭示已成为蛋白质组学的一个重要的研究内容。随着蛋白质组学研究的深入发展，在揭示如生长、发育和代谢调控等生命活动的规律上将会有所突破，对探讨重大疾病的机理、疾病治疗、疾病防治和新药开发将提供重要的理论

基础。对于鉴定药物目标，蛋白质相互作用数据有潜在的价值。几家私人公司也已开始基因组范围的双杂交分析，如Hybrigenics 和 Myriad Genetics Inc(http://www.hybrigenics.com; http://www.myriad pronet.com)，他们都相信理解蛋白质功能和途径更有利于开展药物的研究。

用酵母双杂交及其他方法得到的蛋白质相互作用产生了几个数据库，可以用来构建一个大的相互作用的网络[18]。在全面分析了酿酒酵母中 2709 个蛋白质相互作用后，建立了一个由 1548 个蛋白质间 2358 个相互作用组成的一个巨大的网络和几个较小的网络，这表明细胞过程的复杂性。这个网络指明了不同细胞部位间蛋白质的相互作用，最重要的是提供了未知蛋白质的功能线索[19]。特定功能种类的蛋白质和许多其他类的蛋白质有相互作用，如细胞周期调控、转录和染色体调控的蛋白质。这和它们在细胞中的中心地位是一致的。通过简化这个网络图，24 个核心蛋白中有 21 个和生长控制、染色体结构/染色质重排或转录有关[20]。要指出的是，酵母双杂交得到的蛋白质相互作用是可能发生的相互作用，在真正的细胞环境中不一定发生，因此需要用其他的方法进一步验证。随着技术的革新，双杂交系统的“敏感性”和“专一性”不断提高，高通量和自动化的实验方法可以完成更复杂的蛋白质组中的蛋白质相互作用，在阐明蛋白质的功能、揭示生命现象的规律、弄清重大疾病的机理和新药开发中发挥重要的作用。当前的连锁图的模式缺乏时间和空间上的分辨率，几个分享一个共同蛋白质的不同的复合体可能被人为地互相联系到一起。分类、作图和定位这些相互作用将画出一个真正的功能和动态的蛋白质相互作用图谱，使我们对细胞作为一个分子系统有全面的了解。

四、酵母双杂交技术方法

双杂交筛选的第一步是选择适当的载体系统。有许多 BD 和 AD 载体已被成功的应用，其中应用最广泛的是以 Gal4 为基础的载体。另外有系统利用细菌 LexA 蛋白做 BD，VP16 做 AD。Gal4 和 LexA 系统各有优缺点，有许多已经商业化，在选择时可参考产品说明。选定载体后，将待鉴定蛋白分别构建到 BD 和 AD 载体上。构建载体时，除了常规的克隆方法外，用同源重组的方法更适合于大规模构建重组质粒，这种方法不用繁琐的基因操作。同源重组方法先用 PCR 扩增目的基因，基因两端带有和载体同源的序列，当 PCR 产物和线性化的质粒共转化到酵母细胞后，通过同源重组缺口修复产生重组质粒。据报道，在酵母中 DNA 片段的末端有 30 个碱基的同源区就足以使该片段整合到线性化的质粒中[21,22]。此外有很多已经商业化的构建在 AD 载体上的来自不同生物或不同组织的 cDNA 文库可以获得，用于不同目的的筛选。一个常见的筛选文库的双杂交实验流程图 11-1。

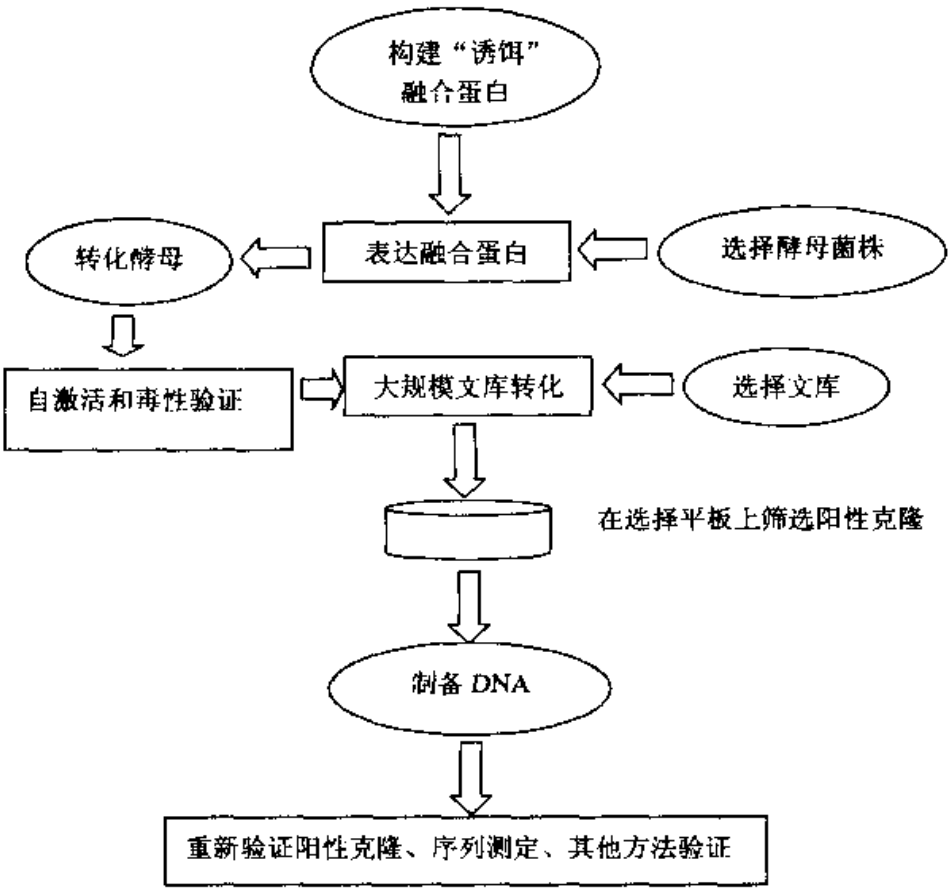


图 11-1

在转入同一酵母细胞前要进行毒性、自激活报道基因和蛋白质表达的检测。比较转化了“诱饵”质粒的细胞和转化空载体的酵母细胞在液体培养基中的生长速率，检测“诱饵”蛋白对宿主的毒性影响。如果“诱饵”细胞生长速率显著降低，“诱饵”蛋白可能是有毒的。因此，在和文库共转化前需要在琼脂平板上生长该酵母细胞株。检测“诱饵”蛋白的转录激活可以将细胞涂于选择性平板上，同时涂带有空载体的菌株于适当的选择培养基上。如果“诱饵”可以自激活报道基因的转录，可以再重新构建“诱饵”质粒，消除转录激活区。另外，如果酵母细胞在缺乏 His 培养基上生长，可以加入适当的 3-AT 消除背景。蛋白质表达的检测可采用 Western 印迹的方法。

构建好的“诱饵”和“猎物”质粒需要共转化到同一宿主细胞中，然后筛选激活报道基因的阳性克隆。在双杂交分析中要经过两次转化，工作量相当大，尤其是在筛选文库的时候更是如此。酵母的转化效率比细菌要低 4 个数量级。因此转化步骤成为双杂交技术的瓶颈。转化的基本思想是得到尽可能多的转化子，同时要避免在同一酵母细胞中转化多种文库质粒，因为这会影响随后的分析。但是用最少量的 DNA 得到最大数量的转化子需要高的转化效率。转化酵母细胞有几

种常用的方法，包括原生质体法、电转化法和醋酸锂转化法。虽然原生质体法能得到高的转化效率，对于大规模的实验如文库筛选时间花费太多。电转化法有最高的转化效率，用小量的 DNA 做转化时非常有效，电转化法只能用小量的 DNA，因而限制了它在双杂交方法中的应用。目前普遍应用的大规模的转化方法都采用醋酸锂转化法[23]，它是一种在大规模实验中快速在酵母细胞中导入 DNA 的方法。最近有文献报道用醋酸锂和 DTT 预处理酵母细胞，然后用电转化的方法共转化质粒。该方法比化学转化方法转化效率要高 18 倍，因此节约了大量的时间和花费，适合于高通量的大规模的筛选相互作用[24]。另外，Bendixen (1994) 等人[25]通过酵母接合型的应用，避免了两次转化，提高了双杂交的效率。在酵母的有性生殖过程中有两种接合型：a 和 α，这两种单倍体之间接合能形成二倍体，但同一接合型间的细胞不能形成二倍体。将“诱饵”和“猎物”质粒分别转化不同接合型的酵母细胞。过夜混合培养形成二倍体，在选择性培养基上筛选发生相互作用的二倍体。

目前用接合方法进行大规模双杂交的方法可以分成两种。一种是阵列筛选法，用不同接合型的表达“猎物”和“诱饵”蛋白的酵母株一一接合。用这种方法可以推测已知的蛋白质间的相互作用，例如从最近测序和注释的基因组中分析全长开放阅读框之间的成对的相互作用。虽然是一对一的方法，由于克隆方法和利用小型机械化的工作站，此方法可用于大规模的研究。另一种方法是文库筛选法，用表达一种“诱饵”蛋白的酵母细胞和一个表达复杂的文库“猎物”蛋白的酵母细胞直接接合。这种高通量的系统可以发现许多未知的蛋白质。然后通过激活报道基因来筛选有相互作用的蛋白质。常用的报道基因有 His3、lacZ、URA3、ADE2 和 MEL1 等。His3 是编码组氨酸合成的基因，3-AT 是它的竞争性抑制剂；lacZ 是编码 β-半乳糖苷酶的基因，可以用 β-Gal 实验测定活性。URA3 的编码尿嘧啶合成酶的基因，可以用于正向和反向筛选。ADE2 是编码腺苷酸合成酶的基因。MEL1 编码 α-半乳糖苷酶，可以降解 X-α-Gal，产生蓝色。用不同的启动子控制多个报道基因，可以大大消除双杂交中由于猎物和特定启动子两侧的序列直接作用以及直接和特定启动子的转录因子相互作用而产生的假阳性[26]。

筛选文库的过程中，转化混合物需要在选择性平板上筛选。得到的阳性克隆用克隆 PCR 的方法直接从阳性克隆中扩增未知片段，然后进行测序。Blast (http://www.ncbi.nlm.nih.gov) 搜索 Genbank、EMBL、或其他数据库，分析其代表的基因。对有意义的相互作用用其他的方法如免疫共沉淀、GST-pull down、荧光共定位等进一步验证，得到蛋白质功能的信息。

参考文献

1 Fields S, Song O. A novel genetic system to detect protein-protein interactions. Nature, 1989, 340: 245～

2 Marsolier M C, Prolean M N, Sentenac A. A RNA polymerase III based two-hybrid system to study RNA polymerase II transcriptional regulators. J Mol Biol, 1997, 268: 243～249

3 Vidal M, Brachmann R K, Fattaey A, Harlow E, Boeke J D. Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. Proc Natl Acad Sci, 1996, 93: 10315～10320

4 Shih H M, Goldman P S, DeMaggio A J, Hollenberg S M, Goodman R H, Hoekstra M F. A positive genetic selection for disruptions protein-protein interactions identification of CREB mutations that prevent association with the coactivation CBP. Proc Natl Acad Sci, 1996, 93: 13896～13901

5 Aronheim A, Zandi E, Hennemann H, Elledge S J, Karin M. Isolation of an AP-1 represssor by a novel method for detecting protein-protein interactions. Molecular and Cellular Biology, 1997, 17: 3094～3102

6 Johnsson N, Varsharsky A. Split ubiquitin as a sensor of protein interactions in vivo. Proc Natl Acad Sci, 1996, 91: 10340～10344

7 Wolf S S, Roder K, Schweizer M. Construction of a reporter plasmid that allows expression libraries to be exploited for the one hybrid system. BioTechniques, 1996, 20: 568～574

8 SenGupta D J, Wichens M, Fields S. Identification of RNAs that bind to a specific protein using the yeast three-hybrid system. RNA, 1999, 5: 596－601

9 Osborne M, Dalton S, Kochan J P. The yeast tri-hybrid system: genetic detection of trans-phosphorylated ITAM SH2 interactions. BioTechnol, 1995, 13: 1474～1478

10 Bartel P L, Roecklein J A, SenGupta D, Fields S. A protein linkage map of *Escherichia coli* bacteriophage T7. Nature Genetics, 1996, 12: 72～76

11 Flajolet M, Rotondo G, Daviet L, Bergametti F, Inchauspe G, Tiollais P, Transy C, Legrain P. A genomic approach of the hepatitis C virus generates a protein interaction map. Gene, 2000, 242: 369～379

12 McCraith S, Holtzman T, Moss B, Fields S. Genome-wide analysis of vaccinia virus protein-protein interactions. Proc Natl Acad Sci, 2000, 97: 4879～4884

13 Rain J C, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J, Schachter V, Chemama Y, Labigne A, Legrain P. The protein-protein interaction map of *Helicobater pylori*. Nature, 2001, 409: 211～215

14 Uetz P, Giot L, Cagney G, Mansfield T A, Judson R S, Knight J R, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg J M. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature, 2000, 403: 623～627

15 Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci, 2001, 98: 4569～4574

16 Drees B L, Sundin B, Brazeau E, Caviston J P, Chen G C, Guo W, Kozminski K G, Lau M W, Moskow J J, Tong A, Schenkman L R, McKenzie A 3rd, Brennwald P, Longtine M, Bi E, Chan C, Novick P, Boone C, Pringle J R, Davis T N, Fields S, Drubin D G. A protein interaction map for cell polarity development. J Cell Biol, 2001, 154: 549～571

17 Walhout A J, Sordella R, Lu X, Hartley J L, Temple G F, Brasch M A, Thierry-Mieg N, Vidal M. Protein interaction mapping in C. elegans using proteins involved in vulval development. Science, 2000, 287: 116～122

18 Xenarios L, Eisenberg D. Protein interaction databases. Current Opinion in Biotechnology, 2001, 12: 334～339

19 Schwikowski B, Uetz P, Fields S. A network of protein-protein interactions in yeast. Nature Biotechnology, 2000, 18: 1257--1261
20 Tucker C L, Gera J, Uetz P. Towards an understanding of complex protein networks. Trends in Cell Biology, 2001, 11 (3): 102--106
21 Ma H, Kunes S, Schatz P J, Botstein D. Plasmid construction by homologous recombination in yeast. Gene, 1987, 58: 201--216
22 Hua S B, Qiu M, Chan E, Zhu L, Luo Y. Minimum length of sequences homology required for in vivo cloning by homologous recombination in yeast. Plasmid, 1997, 38: 91--96
23 Walhout A J, Vidal M. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. Methods, 2001, 24: 297--306
24 Helmuth M, Altrock W, Bockers T M, Gundelfinger E D, Kreutz M R. An eletrotransfection protocol for yeast two-hybrid library screening. Analytical Biochemistry, 2001, 293: 149--152
25 Bendixen C, Gangloff S, Rothstein R. A yeast mating-selection scheme for detection of protein-protein interactions. Nucleic Acids Research, 1994, 22 (9): 1778--1779
26 James P, Halladay J, Craig E A. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. Genetics, 1996, 144: 1425--1436

第二节　免疫共沉淀技术

蛋白质是生物体功能的执行者。蛋白质可以与其他生物大分子如核酸、碳水化合物、脂类、尤其是其他的蛋白质分子发生相互作用。在细胞内，蛋白质-蛋白质相互作用是酶作用于特异底物的手段，还指导了信号传导、细胞分裂、DNA复制、转录起始，决定了蛋白质短暂的和持久的存在形式。在细胞外，由于蛋白-蛋白之间的相互作用，使得细胞之间建立了联系：在一个细胞表面表达的配体与邻近细胞的受体结合，或者分泌的配体蛋白与较远细胞的受体结合。近年来已经发展了许多技术检测蛋白-蛋白之间的相互作用。

从全细胞提取物中共沉淀蛋白的技术是一种很有价值的检测蛋白-蛋白之间物理作用的方法。当使用抗体时，这种技术就称作免疫共沉淀。共沉淀技术可以用来研究各种条件下已知蛋白之间的相互作用，还可以鉴定复合物中的组分。共沉淀技术可以单独使用，也可以与其他检测蛋白-蛋白相互作用的手段如酵母双杂交技术和利用纯化的蛋白研究相互作用的共纯化技术相结合使用。

在免疫共沉淀比较典型的实验体系内，细胞裂解后在非变性条件下制备总蛋白提取物。以一种蛋白的抗体（结合于固相亲和介质）特异地免疫沉淀这种蛋白，然后用第二种蛋白或更多种蛋白的抗体做免疫印迹，检测它们是否被第一种蛋白共沉淀。在这个过程中，设立对照以检测相互作用的特异性是很重要的。下面将详细地介绍这种技术。

一、检测蛋白质的存在

第一步是准备检测共沉淀涉及的两种蛋白。通常用 Western blot 的方法检测

参与免疫共沉淀的蛋白质都表达或存在。只有那些可以在非变性条件下免疫沉淀蛋白的抗体才是可以用的。需要注意的是，商品化的某些抗体不适合做免疫沉淀，因此要事先考虑清楚再选择使用。经常使用的策略是在目的蛋白的N端或C端加上一个短肽或表位抗原（epitope）——我们称之为标签（tag），当然如果不影响功能的话，也可以把标签加到蛋白内部。两种最常用的标签是HA和c-Myc，分别从流感病毒血凝素蛋白和人核蛋白c-Myc的一段多肽而来。其他的抗体如FLAG抗体也很实用。有时可以插入串联重复的标签以提高灵敏度。串联标签的拷贝数可以由1个到3个甚至9个之多。

当要沉淀的蛋白分子质量非常接近免疫球蛋白的重链（50kDa）和轻链（23kDa）时，使用短的标签不容易与IgG区分开，这时可以融合一个蛋白或多肽，增加其相对分子质量，而同时要求这种蛋白或多肽可以与偶联到固相介质的小分子特异结合，如谷胱甘肽S-转移酶（GST，可以与偶联到Sepharose上的谷胱甘肽结合）和麦芽糖结合蛋白（MBP，可以与麦芽糖亲和介质结合）是常用的两种亲和分子。虽然它们所应用的已经不属于免疫的范畴，但都是利用了蛋白与蛋白之间特异亲和的原理。

二、制备总蛋白提取物

免疫共沉淀成功的第二步是制备含有最大量与最大活性目的蛋白的总蛋白提取物。全细胞提取物中总蛋白的收率与活性并不能完全等同于特异目的蛋白的收率与活性，因此在免疫沉淀操作前最好将二者都加以鉴定。收率和活性受多种因素的影响。裂解缓冲液中盐浓度与去污剂的微小变化就可能对蛋白收率与活性带来较大的影响，也可以影响细胞破碎的速度与效率。这两种因素对那些与大分子结构如膜和细胞骨架相结合的低可溶性蛋白来说尤为重要。另外，加入蛋白酶抑制剂以防止蛋白的降解也是必需的。

我们推荐研究人员开始阶段先做小量抽提的预实验，使用不同浓度的盐和非离子去污剂。最先开始使用的基本裂解缓冲液应包含以下成分。

1）基本成分。包括缓冲剂（如50 mmol/L Tris-Cl，pH 7.5）、少量的非离子去污剂［如0.1%（*V*/*V*）Triton X-100］、盐离子（如100 mmol/L NaCl）、还原剂（如1 mmol/L DTT）和稳定剂［如10%（*V*/*V*）甘油］。

2）蛋白酶抑制剂。由各种蛋白酶抑制剂混合而成的“鸡尾片”（cocktail tablet）可以从公司买到商品化的产品。开始阶段使用的浓度可以参考为：chymostatin、pepstatin A、leupeptin、antipain各5μg/ml，PMSF、benzamidine各1 mmol/L。

3）螯合剂。大约15 mmol/L的EGTA通常用来螯合对于金属蛋白酶必需的金属离子。但是由于EGTA同时也会抑制那些需要金属离子的正常酶的

活性，因此，有时反而要加入适量的金属离子，或者用 EDTA 来代替 EGTA。

4）磷酸脂酶抑制剂。如果蛋白质的磷酸化状态对于相互作用比较重要的话，裂解缓冲液里就应当包含有磷酸脂酶抑制剂。这一类的试剂包括 2.5 mmol/L 正钒酸钠、10 mmol/L NaF、10 mmol/L β-甘油磷酸。

以上的配方可以作为裂解缓冲液的起始配方使用，实际操作中，可以作适当修改，比如可以将 NaCl 在 0～500 mmol/L、Trition X-100 在 0%～1%浓度之间调整。全细胞提取物的总蛋白浓度定量可以通过 Bio-Rad 蛋白测定试剂盒、Pierce 公司 BCA 测定试剂盒以及 Brad ford 方法进行。然后再取 25～75μg 总蛋白通过 Western blot 分析确定某种特定蛋白质的浓度。一般来讲，最好通过测定另一种蛋白的存在作为标准化的内对照以及免疫印迹的阳性对照，这种蛋白通常是看家蛋白，如细胞骨架或核糖体蛋白、或者是同一研究途径中已经证实存在的另一种蛋白。

三、相互作用特异性的对照实验

对照设置对于确定抗体及蛋白-蛋白作用的特异性是必需的。当蛋白被标记以不同的标签时，对照是比较容易设立的。我们以两种蛋白 A 和 B 的相互作用来说明这个问题，蛋白 A 和 B 在细胞内有天然存在的形式，它们是没有标签标记的，称作 non-A 和 non-B。我们分别将 A 和 B 标记 Myc 和 HA 标签，称作 Myc-A 和 HA-B。将二者的质粒共转染到宿主细胞内表达，标记的和未标记的蛋白于是共存于细胞内，但是前者的量要稍多于后者。以 Myc 抗体结合 Myc-A，以 protein G-Agarose 结合 Myc 抗体，形成复合物，通过离心与上清得以分离。如果 A 和 B 蛋白存在相互作用，那么 Myc-A 既可以与 HA-B 结合，也可以与 non-B 结合，两种形式的 B 蛋白因此可以被 A 蛋白共沉淀而被包含在上述复合物中。以 HA 抗体对免疫沉淀物进行 Western blot 检测，HA-B 即可被检测出来，而 non-B 则不能与 HA 抗体发生结合，因而显示不出条带。同时标记 Myc 和 HA 的蛋白互为对方的阴性对照组。其他没有 HA 标记的蛋白也检测不出来。通过以上设置，可以确定抗体及相互作用的特异性。

如果抗体不是针对标签而是针对天然形式的蛋白，那么使用缺失此蛋白表达的细胞系是一种上策。当然，前提是存在这样的细胞系，此蛋白的缺失表达对于细胞的生存没有致命的影响。如果上述策略不实用，最通用的办法是用预免疫血清（preimmune serum）或同一物种来源的与蛋白没有特异结合的抗体（比如正常的免疫球蛋白 G 等）预先免疫沉淀，以预清除（pre-clear）来降低背景。但是后两种对照不能完全排除间接结合目的蛋白的可能性。

使用蛋白 A/G-Sepharose（或 Agarose）共沉淀蛋白：当制备完成提取物之

后，接下来就可以免疫共沉淀特异结合的蛋白。下面的方法是一种使用了 protein A/G-Sepharose（Agarose 也可）的标准方案。其他的方法如使用了 GST 偶联谷胱甘肽-Agarose 的方法不在此详述。操作中所使用的缓冲液与微量离心管等都必须在冰浴条件下和低温离心机中进行。

材料

总蛋白提取物
抗体
免疫共沉淀缓冲液
5 mol/L NaCl
protein A/G-Sepharose
2×SDS-PAGE 样品缓冲液
低温离心机
其他 SDS-PAGE 和 Western blot 所需的试剂与设备

步骤

1）制备样品，在冰浴的离心管内进行，含有如下成分：
0.5mg～1 mg 全细胞提取物
1μg 抗体
5 mol/L NaCl 母液稀释至 100 mmol/L NaCl
0.5 ml 免疫共沉淀缓冲液
2）温和地颠倒离心管几次，冰浴 90min（此时间可以根据实际情况延长与缩短），随机翻转离心管使溶液充分混匀，这样的操作可以在特定的旋转混匀器上进行。
3）4℃下以最大转速离心 10min，去除沉淀，将上清转移到一新的离心管。
4）加入 50μl protein A/G-Sepharose 匀浆物（含 25～30μl 柱体积），加入前要混匀 protein A/G-Sepharose 匀浆物。
5）4℃下温和地混匀 30～60min。要避免剧烈的振荡操作。
6）4℃下 1000 r/min 离心 30s，沉淀 protein A/G-Sepharose。
7）以 1ml 免疫共沉淀缓冲液洗涤沉淀 3 次。每一次洗涤，在离心前温和颠倒离心管 3 次即可，离心后小心吸出上清。此步中免疫共沉淀缓冲液中的蛋白酶抑制剂不可省去，尽管其价格比较高。
8）从最后一次的离心产物中吸净上清，小心不要吸走 Sepharose，加入 25μl 2×SDS-PAGE 样品缓冲液。
9）煮沸热变性 5min，离心，吸取上清，SDS-PAGE 分离。
10）等量的样品用于 Western blot 分析，分别以两种蛋白特异的抗体（蛋白本身或其偶联的标签）进行每一种蛋白的检测。以等量的全细胞提取物

为阳性对照。

试剂与溶液配方（使用去离子的双蒸水配制所有试剂）

免疫共沉淀缓冲液：

50 mmol/L Tris－Cl，pH7.5

15 mmol/L EGTA

100 mmol/L NaCl

0.1%（*V*/*V*）Triton X－100

以上配制完成后存于4℃，临用前加入：

1×蛋白酶抑制剂混合物（配方见下）

1 mmol/L DTT

1 mmol/L PMSF（最后加入）

1000×蛋白酶抑制剂混合物（溶于DMSO中）

5 mg/ml chymostatin

5 mg/ml pepstatin A

5 mg/ml leupeptin

5 mg/ml antipain

可分装后在－20℃保存1年。

四、免疫共沉淀技术的其他分析

共沉淀技术是一种有力而又简便的检测蛋白与蛋白之间物理作用的方法。它之所以能用于这类研究，有如下的原因。首先，作为蛋白亲和层析的一种形式，这种方法具有足够的灵敏度来检测弱的相互作用；其次，共沉淀实验在1 mg全细胞提取物内进行，蛋白所处的浓度条件与环境条件都是最接近天然状态的，能够反映体内的相互作用情况，这使得共沉淀技术成为与酵母双杂交相结合使用的很好的一种鉴定蛋白-蛋白相互作用的手段。在酵母双杂交中，当蛋白质结构域由于折叠变化而不正确地暴露时，可以导致假阳性的出现。另外，并非所有的蛋白都能适用于酵母双杂交体系，因此阴性的结果有时可能掩盖了真正的相互作用。共沉淀技术因此也是酵母双杂交技术的一种互补。

然而，需要提醒注意的是，从细胞提取物中共沉淀两种蛋白也不能充分地说明这种特异的相互作用就能在体内发生。还需要其他的实验给以佐证，比如细胞共定位的研究（见本章第三节）和功能相关性分析。

当进行共沉淀实验时，非常重要的一点是从两个方向分别进行共沉淀，即以蛋白A的抗体免疫沉淀，以蛋白B的抗体免疫检测；同时以蛋白B的抗体免疫

沉淀，以蛋白 A 的抗体免疫检测。这样做可以提供更为信服的证据表明二者相互作用的存在是真实的。但是，有时只能从一个方向成功地免疫共沉淀。比如，蛋白结合抗体或其他亲和试剂阻断了同另一种蛋白的结合，或者由于蛋白结合比例的差别造成，例如，蛋白 A 可以结合除蛋白 B 之外的其他许多蛋白，而蛋白 B 则相反，大部分只能与蛋白 A 结合，在这种情况下，最为有效的方法是免疫沉淀蛋白 B，而不是蛋白 A。

蛋白/全细胞裂解物的制备与免疫沉淀的条件可变性比较大，哪种更为有效，要看情况而定。当相互作用检测不到的时候，可以将条件的严谨度放松，比如降低盐浓度，减少甚至去除非离子去污剂。根据相互作用的强度，可以调整缓冲液的组分，如 1% Triton X-100，0.5% 脱氧胆酸钠，0.1% SDS。

此外，有必要提高蛋白的表达水平，使之更容易被检测与共沉淀。表达水平有一个范围，太高也不全是好事，它会使得相互作用变得不受调节。当表达水平不能得到很大提高时，可以增加免疫共沉淀所用的全细胞提取物的量，目的是使其中所包含的目的蛋白的量足以共沉淀[2]。因此，最好是使全细胞提取物的浓度较高，这样使用较小体积就可达到实验需要。当一种蛋白在细胞内内源表达很低时，可以过表达标签标记的蛋白，大规模表达，再通过一定的方法浓缩，使蛋白浓度增大。

免疫共沉淀中最为重要的目标是获得一个信噪比低、背景干净的结果。改变某些参数可以提高免疫共沉淀的成功率。其中之一是优化免疫沉淀所用的抗体。Protein A-Sepharose 和 protein G-Sepharose 可以给出相当于抗 Ig 血清的结果。使用直接偶联有 Sepharose 的抗体用作免疫沉淀，会降低背景。此外，抗体与全细胞提取物的比例及全细胞提取物的用量很大程度上影响了结果的好坏。亲和层析纯化的抗体要比那些未经纯化的抗体更能减少交叉反应。

Protein A/G 对于不同物种来源的抗体具有不同程度的亲和力，可以根据下面的表 11-1 列出的亲和力比较选择合适的种类用于免疫共沉淀。

还有一些其他的降低背景的方法。首先，在 100 000 g 的高转速下离心，可以更好地分离细胞提取物，而且可以直接用于免疫沉淀，而不需经过冷冻，这样可以提高沉淀的效率。其次，裂解缓冲液与免疫共沉淀缓冲液中都可以加入 1% BSA 来降低非特异的结合。第三，全细胞提取物可以与 protein A/G-Sepharose 预先孵育，以去除非特异结合的蛋白。第四，在共沉淀与洗涤时可以增加盐离子与去污剂的量来减少非特异结合。第五，增加洗涤的次数可以去除非特异结合的蛋白，但这样做同时也会减少特异结合的相互作用的蛋白的量。另外，还可以制备目的蛋白更浓的细胞提取物，比如，如果已知目的蛋白为核蛋白，就可以提取核裂解物用于随后的实验。最后，可以选用更容易出好结果的抗体标签标记目的蛋白。

表 11-1　protein A/G 与不同物种抗体的亲和力比较

Antibody	Protein A	Protein G
Human	+ + + +	+ + + +
Horse	+ +	+ + + +
Cow	+ +	+ + + +
Pig	+ + +	+ + +
Sheep	+ / -	+ +
Goat	-	+ +
Rabbit	+ + + +	+ + +
Chicken		+
Hamster	+	+ +
Guinea pig	+ + + +	+ +
Rat	+ / -	+ +
Mouse	+ +	+ +

令人信服的结果应该在至少 3 次独立的实验中稳定地重复出来。

参 考 文 献

1　Elaine A E. Current Protocols in Molecular Biology, 1999, 20. 5.1～20.5.9

2　Feng Y, Song l. Y, Kincaid E, Mahanty S K, Elion E A. Functional binding between G and the LIM domain of Ste5 is required to activate the MEKK Ste11. Cur Biol, 1998, 8: 267～278

第三节　细胞共定位技术

不论是转染到细胞内的质粒过表达的蛋白质产物，还是细胞内源性的基因表达产物，在细胞内都会有特定的定位，随着细胞周期的变化、发育阶段的不同和组织分布的差异，同一种蛋白的定位可能还会发生一定的变化。目前，已经发展了多种方法来确定蛋白质在细胞内的分布。比如细胞不同组分的分布抽提法、运用荧光显微镜直接观察等方法是目前比较适用于实验室内操作的、技术相对比较成熟的两种技术。由于它们在蛋白质-蛋白质相互作用领域，特别是近年来在蛋白质组学研究中的应用越来越广泛，本节将对这一类技术作简要的介绍。

一、利用荧光蛋白融合技术研究蛋白的细胞内定位

绿色荧光蛋白（green fluorescent protein, GFP）由 238 个氨基酸残基组成，

最早从华盛顿星期五港水域的维多利亚水母体内克隆得到[1]。后来发现其他品种的水母也存在类似的荧光蛋白。一些腔肠动物通过体内的荧光素酶氧化荧光素，释放能量，激活 GFP，而水母却通过水母蛋白激活 GFP[2,3]。目前关于 GFP 的认识，主要来自 Prasher 等最早得到的克隆（GFP10.1）及其衍生物[4]。

Chalfie 等人进行了揭示 GFP 作为报告蛋白的潜在应用价值的实验[1]。他们发现大肠杆菌或秀丽隐杆线虫（*C. elegans*）中表达编码 GFP 的 cDNA 都能得到 GFP 的特异荧光。这一证据有力地证明了 GFP 蛋白本身具有发光功能的荧光团，无需其他的辅助因子或特殊蛋白的参与。此后，多家实验室尝试了 GFP 基因在多种系统中的表达，发现 GFP 可以在酵母、果蝇、植物、两栖类、哺乳类动物细胞和大鼠脑神经元细胞中成功地表达[1,5~15]。GFP 是一种独特的报告基因，不需抗体、辅因子、酶底物等其他成分。GFP 不影响宿主细胞，因而可以鉴定、跟踪、分选表达 GFP 的活细胞。

目前已经得到了 GFP 的晶体，其立体结构也得以阐明，在 Brookhaven 国家实验室网址http://www.pdb.bnl.gov/browse_it.的 I. D. 代码为1 ema。GFP 由 11 条链的 β 桶形结构包围一个中央螺旋组成。蛋白上任何一个小的缺损都会破坏其结构。缺失实验表明，从第 2 个氨基酸残基一直到第 232 个残基都是 GFP 形成荧光所必需的[16]。许多突变可以影响 GFP 荧光激发与逸出的特征。

由于细胞本身没有 GFP 基因，将 GFP 和另一基因的编码区相连得到的融合体可以研究目的蛋白的定位。在许多细胞中，GFP 均匀地分布于细胞质、细胞核和诸如轴突的远端突起。激光共聚焦显微术发现，只有囊状小泡没有标记荧光。同时，GFP 也并不是在所有的细胞系中都能如此均匀地分布，例如，在 COS 细胞中，GFP 集中分布在细胞核中。由于存在这样的例外情况，在分析融合蛋白产生的荧光之前，首先研究 GFP 在一个未知的细胞系的定位特点显得非常重要。

美国的 CLONTECH 公司开发了一系列活细胞荧光定位观察的载体，具有如下特点：可以直接检测与观察，确定目的蛋白的定位；不依赖于物种，可以在细菌、酵母、果蝇、线虫、斑马鱼、爪蟾、小鼠、人源的多种细胞以及许多种植物细胞内表达；半衰期超过 24h，在活细胞内，荧光可以持续 24h 以上，在固定的细胞内，荧光在数月后仍可以检测得到。鉴于以上优点，目前国际上许多实验室都是应用 CLONTECH 公司开发的产品。其荧光载体携带的荧光蛋白分为两类：一类是前面提到的绿色荧光蛋白，而且就是最早所发现的维多利亚水母（*Aequoria Victoria*）体内的那种，研究人员对野生型的 GFP 进行了定点突变，得到了三种变异形式，分别称为 EGFP、ECFP、EYFP，当受到紫外线或蓝色光照射的时候，可以激发显示出绿色荧光；另一类是红色荧光蛋白，是 1999 年在印度-太平洋生存的一种名叫 Discosoma 的海生珊瑚虫体内发现的[17]，这种新发现的红色荧光蛋白与 GFP 没有任何同源性，可以在哺乳动物细胞内高表达，编码

227个氨基酸。以上两类载体在设计时分别作了一定的密码子修饰，使之更适合于在哺乳动物细胞内表达。除此之外，CLONTECH还开发了适合在大肠杆菌和酵母细胞内表达的载体，采用lac启动子而非CMV IE启动子。常用的适合哺乳动物细胞体系表达的载体名称是：pEGFP-N1/2/3；pEGFP-C1/2/3；pDsRed1-N1；pDsRed1-C1；pECFP-N1；pECFP-C1；pEYFP-N1；pEYFP-C1。其中N代表多克隆区即外源基因在GFP或RFP的氨基端，C代表多克隆区即外源基因在GFP或RFP的羧基端，N1/2/3分别代表三种读码框架。

利用包含翻译起始位点和终止位点的引物，PCR扩增靶基因，构建重组克隆。在引物中还需引入合适的限制性酶切位点，所谓合适是指存在于载体多克隆区、而在插入的外源基因内部不包含的限制性内切酶。通过PCR技术扩增到两端附加了酶切位点的外源基因片段，与T载体（可以很方便地从PROMEGA等公司购买到商品化的载体）连接，DNA测序确认外源基因的序列未发生突变，而后再从T载体上经过设计的酶切组合切下外源基因，亚克隆到pEGFP-N1等载体中，即完成荧光定位表达质粒载体的构建。

利用荧光定位载体可以很方便地研究两种或两种以上蛋白的定位情况。如果蛋白与蛋白之间存在生理的相互作用，那么它们在空间上可能具有共同的细胞内定位，或者，一种蛋白的表达会影响到另一种蛋白的定位情况。因此，在研究多种蛋白之间是否存在共定位时，每一种蛋白的单一定位是必须要考虑的阴性对照。我们以pEGFP-N1和pDsRed1-N1两种载体为例，解释研究两种蛋白A和B是否存在共定位的过程。

首先，按照前述的方法构建pEGFP-N1-A和pDsRed1-N1-B两种重组质粒；然后通过脂质体介导或其他合适的转染方法，将以上质粒“单独”和“共同”转染真核细胞；24～48h后，在常规的透射荧光显微镜或者共聚焦显微镜下观察。为了确定细胞核的位置，在观察前最好先用冰甲醇在－20℃固定10min，或者用4%多聚甲醛在室温固定20min，而后用Hoechst 33258或33342或者DAPI染细胞核10min，使之与细胞核的DNA/染色质结合，再以PBS洗涤后观察。我们假设蛋白A单独转染时定位于细胞核内，蛋白B单独转染时分布于细胞质和细胞核内，二者共转染后在蓝色光激发下融合GFP的A蛋白定位于细胞核内，显示绿色；在绿色光激发下融合RFP的B蛋白也定位于细胞核内，显示红色，而不再有细胞质的定位，说明A蛋白的表达影响了B蛋白的定位，使之由胞浆、胞核广泛分布变化为特异的细胞核定位，而且B蛋白与A蛋白具有共同的核定位。利用冷CCD照相机可以拍摄同一视野不同激发光下的不同颜色的图像，结合图像分析软件，将绿色、红色以及细胞核-DAPI显示的蓝色等多种图像叠加，更加直观地证明A和B蛋白的共定位以及共同定位于细胞的什么位置。

为了更加确定性地把握一种蛋白质在细胞内的精确定位，仅仅靠Hoechst或DAPI区分细胞核是不够的。CLONTECH公司的研究人员开发了一系列不同亚

细胞成分的定位 marker，可以分别将 GFP、CFP、YFP 或 RFP（分别为绿色、蓝色、黄色或红色）与代表特定亚细胞组分的蛋白融合，表达在细胞的特定区域，作为判断其他蛋白定位的标准。目前可以获得的载体所代表的亚细胞组分及其蛋白质分别是：肌动纤维（微丝）—人肌动蛋白；微管—人微管蛋白；内质网—钙泵蛋白的靶向序列；高尔基体—人 β-1，4-半乳糖基转移酶；膜—c-Ha-Ras 的法尼酰化序列；线粒体—细胞色素 c 氧化酶的亚基Ⅷ的靶向序列；细胞核—SV40 T-抗原的核定位信号；过氧化物酶体—过氧化物酶体靶向信号 1。

在另外一些情况下，我们可能会观察到两种蛋白单独表达与共同表达的定位结果相同，而且不具有共同的细胞内分布，比如 A 蛋白分布在细胞质，B 蛋白分布在细胞核，定位具有明显的差异。这个时候我们还是无法确定地说 A 蛋白与 B 蛋白就不存在相互作用，需要考虑的因素有细胞类型、细胞周期、细胞的生理状态、所受到的外界信号刺激等。某些蛋白的定位强烈依赖于细胞类型，在一种细胞系内可能定位在细胞质，在另一种细胞系内可能定位在细胞核，需要具体情况具体分析。很多的研究结果显示，蛋白质的磷酸化和去磷酸化在很大程度上影响了蛋白的定位。例如，在 HER2/NEU 过表达的细胞系内，p21 蛋白在胞浆完成蛋白质的合成之后，可以进入细胞核内，调控细胞周期的变化。但是，当 p21 被丝氨酸/苏氨酸激酶 Akt/PKB 磷酸化之后，就不能进入细胞核，而被阻滞于细胞质内。因此，共定位的研究要充分考虑到多方面的复杂的细胞内因素，尤其是在开展蛋白质组学方面的研究时，特异的组织与发育阶段，可能使我们得出的结论不完全相同，甚至相反。

GFP 融合蛋白的方法在操作上比较容易掌握，步骤简便，可以随时观察，根据荧光表达的强弱适时选择观察时间。但是这种方法也有其限制之处，由于绝大多数细胞内没有内源性的 GFP 表达，因此靶蛋白与 GFP 的融合蛋白必须经过转染，在细胞内过表达后，才能观察。由于过表达的蛋白在细胞内的分布有可能同生理条件下的自然分布有差异，因此建议在过表达实验进行之后，再通过下面将要谈到的免疫荧光技术来确定内源表达蛋白的定位。

二、免疫荧光法研究蛋白-蛋白之间的细胞共定位

免疫荧光染色的基本原理是将已知的抗体或抗原分子标记上荧光素，再与其相对应的抗原或抗体起反应，从而形成的免疫复合物上带有一定量的荧光素，在荧光显微镜下就可以看见发出荧光的抗原抗体结合部位，检测出抗原或抗体。用已知的荧光素标记抗体检测相应抗原的方法称荧光抗体法，用已知的荧光素标记抗原检测相应抗体的方法称为荧光抗原法。

按照抗原抗体反应的结合步骤，免疫荧光染色方法可以分为以下三种。

1. 直接法

用荧光素标记的特异性抗体直接与相应抗原结合，以检查出相应的抗原成分。

2. 间接法

检测未知抗原时先用特异性抗体与相应抗原结合，洗去未结合的抗体，再用荧光素标记的抗特异性抗体（间接荧光抗体）与特异性抗体相结合，形成抗原-特异性抗体-间接荧光抗体的复合物。在此复合物上带有比直接法更多的荧光标记物，所以，此法比直接法敏感。检查未知抗体时，先用已知抗原与细胞或组织中的抗体反应，再与特异性荧光抗体反应形成抗体-抗原-特异性荧光抗体复合物。或用固相已知抗原检查液相（如血清）中的抗体，形成抗原-抗体-间接荧光抗体复合物，从而将抗体检查出来，称夹心法。

3. 补体法

用特异性抗体和补体的混合液与标本上的抗原反应，补体就结合在抗原抗体复合物上，再用抗补体的荧光抗体与补体结合，从而形成抗原-抗体-补体-抗补体荧光抗体复合物。荧光显微镜下所见阳性荧光即为抗原所在部位。补体法不仅敏感，而且不受特异性抗体种属的限制，适用于各种不同种动物抗体的检查方法。

在研究蛋白质与蛋白质相互作用的实验中，最常用的免疫荧光技术是间接法。抗原即是我们所要研究的靶蛋白。有些情况下，我们可以将靶基因构建到商品化的表达载体上，因此可以选择的特异性抗体既可以是抗表达载体上的标签的商品化抗体（比如，CLONTECH 公司开发的 pCMV-Myc 载体和 Invitrogen 公司开发的 pcDNA3.1-Myc/His 表达载体都设计有 Myc 表位抗原，可以同 Myc 抗体特异结合），也可以是抗靶蛋白的特异性抗体（如果没有商品化的抗体，需要制备足量抗原，免疫动物，制备抗体），增加了操作的灵活性。这种抗体我们将之称为一抗。所谓的间接荧光抗体，是偶联有荧光素的二抗，二抗具有物种特异性，但不具有抗原特异性，因此，一种二抗可以应用于许多种一抗。而且经过一抗、二抗两级结合，可以将信号放大，因此间接法比直接法更为灵敏，尽管在操作上稍微复杂一些。

可通过选择不同的荧光素种类，分别标记不同的蛋白，研究同一个体系内不同种蛋白的定位情况。最常用的荧光素有 FITC 和 TRITC、Texas red、Rhodamine。前者具有类似 GFP 的观察效果，可以发绿色荧光；后二者具有类似 RFP 的观察效果，可以发红色荧光。

下面的程序是一个具体的例子，以说明这种方法的操作。

实验目的：检测 GSK-3β 与微管蛋白（β-tubulin Ⅲ）是否具有共定位；

操作程序：

1）细胞培养：人成神经细胞瘤细胞系 SH-SY5Y 按常规方法培养在六孔板或盖玻片上。

2）固定：准备4%多聚甲醛和（或）100%冰甲醇，以 PBS/TBS 洗涤细胞一次，吸净 PBS/TBS，立即固定，可以选择如下两种方法中的一种：

甲醇固定法（蛋白质沉淀法）：

a. 将细胞在 100%冰甲醇中于 -20°C 固定 10min；

b. 室温下以 TBS 洗涤盖玻片 3 次，每次 5min；

c. 以 TBS 新鲜配制的 0.1% 硼氢化钠猝灭细胞 5min；

d. 室温下以 TBS 洗涤 3 次，每次 5min，吸净 TBS，将盖玻片浸在 TBS 内，继续下面的封闭步骤。

多聚甲醛固定法（交联法）：

a. 将细胞在 4% 多聚甲醛中于室温固定 10min；

b. 以 TBS 洗涤盖玻片一次，吸净 TBS，以 0.2% Triton X-100 室温处理 5min 或 100%冰甲醇中于 -20°C 处理 5min；

c. 室温下以 TBS 洗涤 3 次，每次 5min；

d. 以 TBS 新鲜配制的 0.1% 硼氢化钠猝灭细胞 5min，吸净 TBS，将盖玻片浸在 TBS 内，继续下面的封闭步骤。

3）封闭：

在封闭液（10% 马或羊血清，1% BSA，0.02% NaN_3，1×PBS）中室温封闭 40～60min，而后以 PBS 洗涤一次，5min；

4）加入一抗——小鼠抗 GSK-3β 单克隆抗体（稀释到含 2% BSA 的 PBS 中），抗体终浓度为 10μg/ml，4℃孵育过夜。（注意：当第一次使用某种抗体时，要先做预实验确定最合适的抗体稀释比例，使得结果特异而且背景最低。）

5）以 PBS 洗涤 3 次，每次 5min；

6）加入 FITC 偶联的抗小鼠抗体（二抗，稀释到含 2% BSA 的 PBS 中），二抗终浓度为 10μg/ml，室温孵育 1h（避光）；

7）另外一组盖玻片，加入小鼠抗 β-tubulin Ⅲ单克隆抗体（稀释到含 2% BSA 的 PBS 中），抗体终浓度为 21μg/ml，室温孵育 1h；

8）以 PBS 洗涤 3 次，每次 5min；

9）加入 Texas red 一偶联的抗小鼠抗体（二抗，稀释到含 2% BSA 的 PBS 中），Texas red 终浓度为 5μg/ml，室温孵育 1h；

10）以上两组样品以 PBS 洗涤 1 次，去离子水洗涤 2 次，而后孵育在 50ng/ml DAPI（以水溶解）中，染细胞核，室温下 1h；

11）以去离子水漂洗，置于载玻片上，荧光显微镜下观察，绿色荧光下观察到

的是 GSK-3β 的定位，而红色荧光下观察到的是 β－tubulin Ⅲ的定位。

免疫荧光与 GFP 融合蛋白技术相比，具有可以检测内源性蛋白表达的优势，因此可以反应生理条件下的蛋白质定位情况。由于两种技术具有互补性，所以常常在实践中结合使用。比如，其中一种蛋白可以通过免疫荧光技术来确定定位，而另外一种则通过 GFP 融合技术来确定分布。具体如何，要充分发挥和利用自己实验室的条件，合理选择适合自己的方法。

三、抽提细胞不同组分的方法

前面介绍的两种方法都依赖于荧光显微镜的使用，便于直接观察，可以非常直观地说明不同蛋白质的共定位与否。但是由于荧光显微镜或激光共聚焦显微镜是比较昂贵的仪器，在有的实验室可能不具备使用条件。这时可以运用抽提细胞不同亚细胞组分的方法来分步分离蛋白质，然后通过免疫印迹等方法确认所感兴趣蛋白的存在。

在过去的 40～50 年中，细胞的分级分离为生物学家了解细胞器和生物大分子的组成和功能提供了可能。最近，成功地制备了分级的细胞匀浆物（以无细胞体系或体外系统而为大家所知），这样就为分子生物学家开辟了一个新的舞台。

细胞的分级分离过程包括两个主要的阶段：破碎细胞，然后是细胞成分的分离。破碎培养细胞或组织细胞可以采用渗透压冲击、超声波振荡、机械力研磨或剪切等各种方法来实现。比较温和的破碎程序可以使细胞核、高尔基体、线粒体和其他膜结构保持完整。单层贴壁生长的细胞培养物可以在蔗糖等渗缓冲液（通常为含有 0.25mol/L 蔗糖、补加了单价或二价阳离子、蛋白酶抑制剂和螯合剂等的等渗缓冲液）中匀浆。通常来说，细胞匀浆物分步分离的第一步是差速离心，在一系列离心过程中离心力越来越大。典型的差速离心顺序在表 11-2 列出。依次差速离心得到的沉淀再进行超速离心。这种逐渐加大离心力的顺序离心的过程将根据细胞成分的大小和密度来实现分离。

表 11-2　差速离心形成的沉淀（以肝脏为例）

沉淀	RCF（gav）×时间	内容物
P1	1000g × 10min	细胞核，重线粒体，大片细胞膜
P2	3000g × 10min	重线粒体，细胞膜碎片
P3	6000g × 10min	线粒体，溶酶体，过氧化物酶体，完整高尔基体
P4	10 000g × 10min	线粒体，溶酶体，过氧化物酶体，高尔基膜
P5	20 000g × 10min	溶酶体，过氧化物酶体，高尔基膜，粗面内质网
P6	100 000g × 10min	从内质网而来的所有小泡，细胞膜，核内体等

（引自 Dealtry, G. B. and D. Rickwood, eds. 1992. Cell Biology labfax. BIOS Scientific Publishers, Oxford, United Kingdom.）

在蛋白质与蛋白质相互作用的研究中，涉及最多的是胞浆蛋白与胞核蛋白的鉴定，因此本节所涉及的细胞组分抽提技术主要介绍从细胞培养物中分离细胞核蛋白的方法。其他亚细胞组分的提取方法这里不再赘述（见本书其他章节）。

第一步，细胞裂解。贴壁细胞转染 48h 后，去除 DMEM 培养液，加入预冷却的 PBS，洗涤细胞两次，用细胞刮将细胞小心刮下，4℃离心，1500g，5min，收集细胞沉淀，然后加入 50～100μl 匀浆缓冲液（配方见下），冰浴 10min，再加入 1/10～1/5 体积的 75% 蔗糖，混匀。12 000g，离心 10min，去除上清。

第二步，核裂解。取沉淀，加入 50～100μl 核提取缓冲液（配方见下），并于 4℃振荡 30min。12 000g，离心 30min，取上清，即为核蛋白溶液。

匀浆缓冲液配方：

10 mmol/L Hepes，pH 7.8

0.75 mmol/L spermidine

0.15 mmol/L spermine

10 mmol/L KCl

1 mmol/L DTT

0.1 mmol/L EDTA

0.1 mmol/L EGTA

0.5 mmol/L PMSF

临用前补加：

1 mmol/L NaF

1 mmol/L Na_3VO_4

1 mmol/L $Na_4P_2O_7$

10 μg/ml aprotinin

10 μg/ml pepstatin

10 μg/ml leupeptin

核提取缓冲液配方：

50 mmol/L Tris-Cl pH7.5

420 mmol/L KCl

5 mmol/L $MgCl_2$

0.1 mmol/L EDTA

0.2 mmol/L DTT

10% sucrose

20% glycerol

临用前补加：

0.1 mmol/L PMSF

细胞共定位技术与酵母双杂交、GST pull-down、免疫共沉淀技术等一起构成了目前国际上公认的鉴定蛋白-蛋白相互作用的几种常用方法，也是目前国内实验室应用较多、技术较为成熟的研究方法，是功能基因组学时代研究基因功能不可或缺的手段，因此，在蛋白质组学日益蓬勃发展的今天，仍然有必要掌握这些技术方法的要领。

参考文献

1 Chalfie M, Tu Y, Euskirchen G, Ward W W, Prasher D C. Green fluorescent proten as a marker for gene expression. Science, 1994, 263: 802～805

2 Chalfie M. Green fluorescent protein. Photochem Photobiol, 1995, 62: 651～656

3 Prasher D C. Using GFP to see the light. Trends Genet, 1995, 11: 320～323

4 Prasher D C, Echenrode V K, Ward W W, Prendergast F G, Cormier M J. Primary structure of the *Aequorea victoria* green fluorescent protein. Gene, 1992, 111: 229～233

5 Flach J, Bossie M, Vogel J, Corbett A, Jinks T, Willins D A, Silver P A. A yeast RNA-binding protein shuttles between the nuvleus and the cytoplasm. Mol Cell Biol, 1994, 14: 8399～8407

6 Nabeshima K, Kurooka H, Takeuchi M, Kinoshita K, Nakaseko Y, Yanagida M. p93dis1, which is required for sister chromatid separation, is a novel microtubule and spindle pole body-associating protein phosphorylated at the Cdc2 target sites. Genes Dev, 1995, 9: 1572～1585

7 Maniak M, Rauchenberger R, Albrecht R, Murphy J, Gerisch G. Coronin involved in phagocytosis: Dynamics of particle-induced relocalization visualized by a green fluorescent protein tag. Cell, 1995, 83: 915～924

8 Baulcombe D C, Chapman S, Cruz SS. Jellyfish green fluorescent protein as a reporter for virus infections. Plant J, 1995, 7: 1045～1053

9 Wang S, Hazelrigg T. Implications for bcd mRNA localization from spatial distribution of exu protein in *Drosophila* oogenesis. Nature, 1994, 369: 400～403

10 Barthmaier P, Fyrberg E. Monitoring development and pathology of *Drosophila* indirect flight muscles using green fluorescent protein. Dev Biol, 1995, 169: 770～774

11 Tannahill D, Bray S, Harris W A. A *Drosophila* E (spl) gene is "neurogenic" in *Xenopus*: A green fluorescent protein study. Dev Biol, 1995, 168: 694～697

12 Wu G Y, Zou D J, Koothan T, Cline H T. Infection of frog neurons with vaccinia virus permits in vivo expression of foreign protein. Neuron, 1995, 14: 681～684

13 Marshall J, Molloy R, Moss G W, Howe J R, Hughes T E. The jellyfish green fluorescent protein: A new tool for studying ion channel expression and function. Neuron, 1995, 14: 211～215

14 Misteli T, Caceres J F, Spector D L. The dynamics of a pre-mRNA splicing factor in living cells. Nature, 1997, 387: 523～527

15 Lo D, McAllister A, Katz L. Neuronal transfection in brain slices using particle-mediated gene transfer. Neuron, 1994, 13: 1263～1268

16 Dopf J Horiagon T M. Deletion mapping of the *Aequorea Victoria* green fluorescent protein. Gene, 1996, 173: 39～44

17 Matz M V, Fradkov A F, Labas Y A, Savitsky A P, Zaraisky A G, Markelov M L, Lukyanov S A. Fluorescent proteins from nonbioluminescent Anthozoa species. Nat Biotechnol, 1999, 17: 969～973

（王　建　张令强）

第十二章　多能干细胞定向分化的蛋白质组研究

一、多能干细胞

干细胞（stem cells）在延长自我更新和分化之间具有选择能力。这种命运选择由内部信号和外部微环境紧密调控，其间各种机制未明。在许多成熟哺乳动物组织内都可鉴定出干细胞，如：上皮组织、血液和种系，干细胞通过分化来补充因正常细胞的老化和损伤丢失的细胞。干细胞也在成人器官中出现，如脑、胰腺，但只表现出有限的再生和更新。尽管干细胞在成人组织中有较强的“塑型”能力，它只能形成一种有限的细胞类型。与之相反，多能干细胞具有形成各种细胞的潜力[1,2]。

未分化的多能干细胞能够发育形成内胚层、外胚层和中胚层细胞。具有自我更新和分化潜力的三种类型多能干细胞系来源于小鼠的胚胎阶段：① 胚胎干细胞（embryonic stem cells，ES 细胞），20 年前首次从小鼠的 8 细胞囊胚的培养中分离出来[3,4]。从那时起，人们企图从其他的种属中产生具有相似生物学特征的细胞系。尽管这些细胞的某些功能可以复制，但最好的模式细胞仍是来源于小鼠[5]。② 胚胎种系细胞（embryonic germ cells，EG），从原生殖细胞中分离出来，即 9.5～12.5d 胚胎期的种系细胞前体。③胚胎瘤细胞（embryonic carcinoma，EC），是由畸胎瘤细胞中干细胞成分建立。所有这三种类型的细胞都具有相同的特征，如：碱性磷酸酶、胚胎抗原 SSEA-1 和转录因子 Oct-4 的表达、细胞周期的短 G1 期和端粒酶的高活性。ES 和 EG 细胞的增殖由 IL-6 家族调控，而 EC 细胞来源于恶性畸胎瘤，因此并不依赖于生长因子和细胞因子，但必须由化学物质来诱导。

在体内，当 ES 和 EG 细胞被注射入小鼠囊胚内，则参与胚胎的发育，导致嵌合体的产生。通过同源重组进行基因打靶之后，ES 细胞可广泛应用于小鼠的“基因剔除”。为研究外源性基因对胚胎发育的组成性表达的影响，可进行“基因插入”的分析。在体外，当 ES 和 EG 细胞聚集培养形成胚胎小体（embryonic body，EB）时，可自发产生分化。因此，这种培养体系被作为 ES 细胞的体外分化模型，用来研究细胞的分化和基因表达谱[6]。

二、胚胎干细胞

ES细胞是一种优化的细胞资源，由于端粒酶的存在，使其在一定培养条件下能保持未分化状态，这种具有无限增殖能力的多能干细胞可广泛用于各种发育相关的分析。人类和鼠类ES细胞都具有分化成组织特异性的干/祖细胞的潜能。然而，对于每种组织和区域性的细胞种类而言，在发育过程中所要求的系列生长因子和细胞因子、级联放大的转录因子和必须与成熟细胞表型协调的相关基因产物是不同的。实际上，ES细胞可以产生任何细胞类型，无论在体内，还是在体外，因此可达到全能细胞的标准[7]。

在未分化状态，ES细胞表达碱性磷酸酶和系列表面糖类抗原表位。不同阶段特异性的胚胎抗原（SSEA）可明确人（SSEA-3、4）和鼠类（SSEA-1）ES细胞所处的阶段[7]。在多能细胞群中，干细胞的维持需要转录因子Oct-4的表达。当ES细胞被触发产生分化时，Oct-4表达下调。体内突变显示：在囊胚阶段Oct-4的丢失可导致内细胞团分化为外滋养层细胞。近来的研究表明，大约50%～150%的内源性表达水平才能允许进行全能干细胞的自我更新和维持[8]。目前，已发现Oct-4呈现一种种属间相似的行为规律[10]。急性时相反应因子STAT3，被几种分化相关的细胞因子亚家族所激活，可产生与Oct-4行为相关的多能干细胞的维持。白血病抑制因子（Leukemia inhibitor factor，LIF）相关的细胞因子似乎是小鼠ES细胞的原型有丝分裂原。从ES细胞培养介质中撤除LIF将导致Oct-4的表达降低，同时启动ES细胞的分化[7]。

ES细胞在有饲养层细胞或LIF等分化抑制因子存在的条件下，能保持未分化状态。撤除LIF，ES自发分化，形成多细胞聚合体，即EB。EB的发育重演了卵裂的过程，EB产生的细胞群分化为原始外胚层、内脏内胚层、早期中胚层，以及许多类型的细胞，如造血干细胞、上皮细胞、心源性、肌源性、软骨源性细胞及神经细胞、内皮细胞和血管平滑肌细胞[10]，这些事件的发生顺序和体内卵黄囊的形成与分化基本相似，表明这种体外模型系统遵循着体内胚胎发育的规则[11]。检测分化中EB表达的细胞系标志基因，与小鼠胚胎着床后发育时期表达的基因相比，发现体外表达的这些基因无论从时间还是从空间顺序来说都与体内基因表达时空模式相符合[6,12]。尽管由于空间控制信号的缺乏，在EB中的形态发育是不可能的，但是，细胞分化仍能以一种可调节的发育和时间依赖的方式来完成。因此，ES细胞分化模型适合于在细胞水平分析发育过程，将成为研究控制复杂系列发育过程分子机制的有益工具。

虽然EB分化过程中基因表达模式与早期胚胎发育很相似，但在EB中，神经细胞及肌细胞周围内胚层和中胚层产物增加，说明在这些系统正常发育中所需要的因子在EB中表达量下降或缺失[13]。分化环境的不同能极大地影响EB形成

不同细胞类型的数量。如在没有诱导者存在的情况下，EB 最终分化的细胞中神经细胞只占极少的比例[14]，要获得大量的神经细胞需要加入维甲酸（retinoic acid，RA）等因子刺激，RA 可促进神经细胞特异性基因的表达，抑制中胚层基因的表达[15]。

目前，由 ES 细胞向神经细胞分化的路线已基本确立。将 ES 细胞转入不黏附底物之中悬浮培养可以得到 EB。EB 继续分化形成包括神经细胞在内的多种细胞混合物，如果用维甲酸诱导可以增加向神经细胞分化的比例。ES 细胞分化而来的神经元中基因表达模式与体内神经细胞分化过程中的基因表达模式极为相似[10]。并且 ES 细胞分化形成的神经细胞和胶质细胞表型正常，植入体内也能行使正常的生理功能[16,17]。这为研究细胞因子、生长因子及激素对神经元发育极早期阶段的作用提供了一个极有效的体外模型。利用此模型，运用同源重组、可诱导的转基因表达、基因诱捕等手段还可进一步分析神经发生及分化的分子调控。

三、人类胚胎干细胞

ES 细胞对于研究哺乳动物发育生物学的巨大帮助已经由对小鼠的大量研究得以证实。尽管哺乳动物的早期发育有很多相似及相同点，但对于进一步研究人胚发育、将 ES 细胞的研究与利用适用于临床，动物模型已远不足用。例如：人类和小鼠的胚盘及所有胚外膜都基本上不同。非人灵长类的 ES 细胞系也许能提供一种研究人类组织分化的较精细的体外模型，但由于一些伦理及实际技术上的原因，包括人类在内的许多灵长类的 ES 细胞不能在嵌合体内形成生殖系以供实验研究。

人类 ES 细胞来源于移植前的胚胎，不同于从其他多能干细胞系，即人类 EG 和 EC 细胞。人类多能 EC 细胞是来源于自发出现的种系细胞肿瘤的非分化的干细胞组分，人类多能 EG 细胞来源于小鼠和人类的原始生殖细胞[18,19]，而人类 ES 细胞来源于囊胚阶段的内细胞团（ICM）。断裂期的人胚，由临床医疗研究的体外受精过程产生。胚胎长至囊胚期后，ICM 被分离，并在被有丝分裂原灭活的小鼠胚胎成纤维饲养层上进行培养。它在含有血清的培养基中可旺盛生长，此时，选择合适的未分化克隆，可进行进一步扩增。从最初的血清中衍生出来后，人类 ES 细胞可以在只含有血清或血清替代介质和碱性成纤维生长因子的饲养层细胞中生长（图 12-1）[1]。

人类 ES 细胞与小鼠 ES 细胞有许多不同。首先，人类 ES 细胞可形成平铺的、紧密的克隆，在胰酶或无 Ca^{2+}、Mg^{2+} 介质中容易形成单个细胞。其次，人类 ES 细胞较小鼠 ES 细胞生长缓慢；小鼠 ES 细胞的倍增时间是 12h，而人类 ES 细胞倍增的时间大约为 36h[20]。第三，人类 ES 细胞的表面抗原表型不同。它表

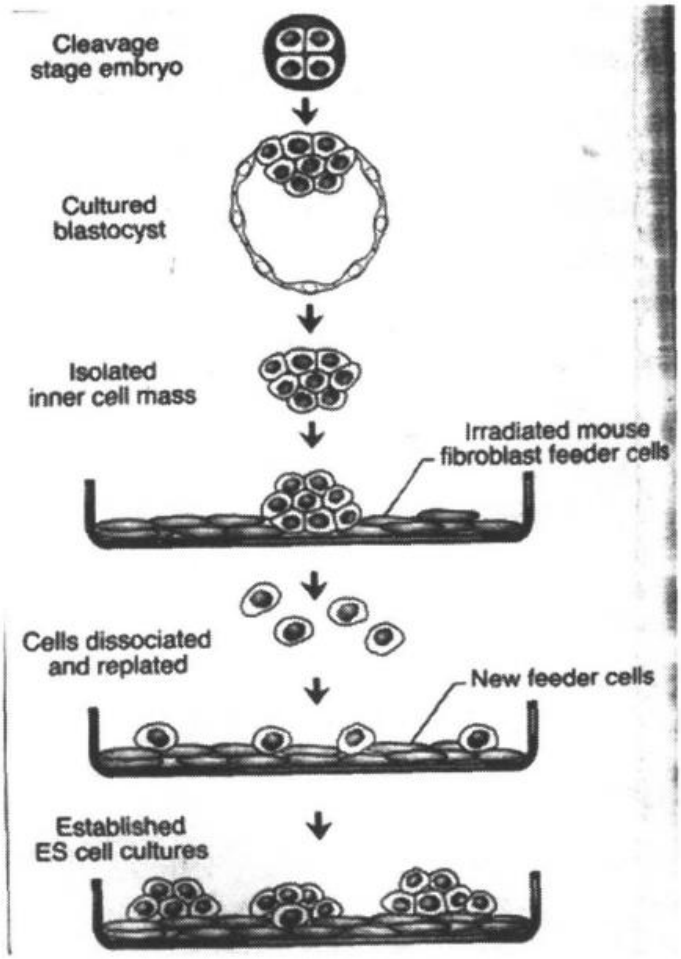


图 12-1　人类 ES 细胞的衍生[1]

达阶段特异性的胚胎抗原 3 或 4（SSEA-3，4）、高相对分子质量糖蛋白 TRA-1-60、TRA-1-81 和碱性磷酸酶。而 SSEA-1 在人类 ES 细胞中不表达[21]。第四，在体外培养方面要求各异。小鼠 ES 细胞要求白血病抑制因子（LIF）以保持其未分化状态；与之相反，LIF 不足以防止人类 ES 细胞的分化[18]，它要求饲养层细胞或者血清的存在，或者存在 bFGF 的无血清培养基[20]。第五，人类 ES 细胞具有明显稳定的染色体核型，这种特征使得人类 ES 细胞更容易成为研究发育的机制模型和移植治疗的已分化细胞的衍生化研究模型。第六，人类 ES 细胞表达高水平的端粒酶，它能够增加端粒重复序列的核糖核蛋白，因此可以维持其长度，与人类细胞系的永生化相关，甚至在超过 300 次倍增时间时，人类 ES 细胞仍显示高端粒酶活性[22]。进一步的工作要求澄清涉及维持人类 ES 细胞增殖的机制、特异性的受体-配体相互作用、下游的信号事件和细胞靶分子。最终，建立不依赖饲养层细胞的培养条件，可允许大规模扩增人类 ES 细胞。

人们一直致力于研究能够在体外培养的人类 ES 细胞，如同其他哺乳动物的 ES 细胞，能够自我增殖，并在一定条件下分化形成多种细胞与组织类型以供基

础研究及移植治疗使用。1998 年有两个实验组分别报道分离出了人类 ES 细胞/EG 细胞[18,22]，它们表现出体外不分化的增殖能力及向多种细胞系分化（三个胚层）的潜能[23~25]。从人类 ES 细胞系衍生出的人类 EB、H9 可以转录 α-fetoprotein、神经纤维 68kDa 亚单位、ζ-珠蛋白和标记原始内胚层的 α-心肌动蛋白、神经外胚层和中胚层衍生物[24]。已分化的细胞可获得与神经元和心肌细胞相似的形态学特征。Odorico JS 等在人类 ES 细胞的体外分化实验中观察到几种细胞形态，包括节律性收缩的心肌细胞、色素或非色素的上皮细胞、显示轴突和树突的神经细胞。Schuldiner 等补充了这些数据，表明在未分化和分化的 ES 细胞中都能表达各种可溶性生长因子的受体，这些生长因子是体内发育通路的效应分子[25]。在利用人类 ES 细胞的其他研究中，Odorico JS 等利用共培养的方法促进了造血干细胞的分化。当 H9、H1 或人类 ES 细胞系的 H1 克隆与非放射性的小鼠骨髓基质细胞一起在未含其他外源性生长因子的胎牛血清中共培养后，可以分化成各种细胞类型。如：表现为典型的造血干细胞前体，可以通过 FACS 鉴定出 $CD34^+$ 细胞；而其他的造血祖细胞克隆也被鉴定出来，如：红系、粒系、巨核系和巨噬细胞系。还可以检测到某些内胚层和胰腺限制性基因，肝细胞核因子 3β，某些胰岛基因，如胰岛素、抑生长素等[1]。总之，这些数据表明人类 ES 细胞在体外能够激活胚胎基因表达程序，并开始向三个胚层的衍生物分化。

将 ES 细胞注射入宿主的囊胚或者移植入免疫相容性小鼠体内，也能产生由多种组织构成的畸胎瘤，证实它们确实是人类的 ES 细胞（EG 细胞），这种体内环境的分化能够揭示未分化 ES 细胞的全部发育趋势。在体内环境下，许多组织构建的正常特征被重现。例如：上皮细胞显示出极性，被一层基膜包裹，同时围绕中胚层；复杂的组织，如毛发、牙齿和胃。在分化的人类 ES 细胞畸胎瘤中鉴定出平滑肌、横纹肌、骨骼、软骨、胃、呼吸上皮、皮肤的角质层、头发、神经上皮细胞和胶质细胞。与人类 EC 细胞相比，人类 ES 细胞可以表现出高级的、均匀一致的发育潜力。由于人类 ES 细胞在体内分化较体外更完全，因此从组织各向异性的混合物中提取有意义的细胞或组织是很有用的。可能达到这个目标的方法有如下几种：① 加入特殊的化学形态原或生长因子；② 将诱导组织或细胞与 ES 细胞共培养或共移植；③ 将 ES 细胞移植入动物的特殊区域或器官；④ 过表达组织特异性的同源盒；⑤ 选择可以激活特殊谱系的基因表达；⑥ 在荧光激活筛选分离细胞的基础上分离有意义的细胞[1]。这些工作是非常令人振奋的，由此不仅掀起了新一轮研究 ES 细胞的热潮，还引起更多的人关心人类克隆及人类胚胎应用等有关伦理道德问题。

由于人类 ES 细胞的建系成功，使得将 ES 细胞应用于临床的研究产生了重要性突破，因此，美国 NIH 委员会在非规定申请资助时间内破例为 ES 细胞的研究提供基金。尽管人类 ES 细胞的一些特性还需进一步的实验证实，而有的特性如形成嵌合体由于伦理的原因尚无法证实，但一种能够自我更新、组织培养下分

化为多种细胞类型的人类细胞必将在基础研究及移植治疗上有着广泛的应用。人们憧憬能够在体外产生特定的细胞、组织、乃至器官，在某种程度上说，这将使人类具有了像低等生物一样的再生能力。人们希望能控制 ES 细胞产生特定的细胞或组织供移植治疗，因为许多疾病仅需一种或少数几种细胞的移植。例如将多巴胺能神经元移植入帕金森氏病人体内、用心肌细胞置换受损的心脏、或用产生胰岛素的细胞治疗糖尿病人。至今为止，小鼠 ES 细胞来源的多种细胞都成功地移植入受体体内，但由于实验条件有限，或限于受体动物种属的年龄，类似实验的长期结果分析还很少，但已说明这种细胞移植的治疗是可能的。尽管首先要克服的是供受体免疫不相容性及免疫排斥反应，但是可以考虑建一个 ES 细胞库，以提供多种不同组织免疫相容性抗原型的 ES 细胞。对 ES 细胞/EG 细胞另一重大利用是发现一些有治疗功能的基因。这些基因可能编码一些具有直接治疗作用的蛋白，如生长因子，或一些药物治疗的重要靶点。利用 ES 细胞来检测新药功能，可以避免新药用于人类临床治疗的一些伦理及法律上的障碍。利用 ES 细胞还可以研究基因功能，它不仅避免了用基因剔除技术所带来的精力、资金和时间上的消耗，还可以研究除基因缺失突变（gene knock-out）外的基因功能获得性突变（gene knock-in），即通过体内发育瞬间的表达或某些时期的长期表达，来研究基因在人类胚胎发育中的作用。

四、维甲酸诱导小鼠 ES 细胞向神经方向分化

类维生素 A（retinol），是一组疏水化合物，包括视黄醛（维生素 A）、维甲酸（RA）及一系列天然或合成的衍生物。在脊椎动物发育过程中，从早期的胚胎发生到维持成体动物上皮组织正常的分化，类维生素 A 都行使重要的生物功能[26,27]。类维生素 A 既能促进也可抑制细胞的生长[28,29]。在许多靶组织中视黄醛（维生素 A）经氧化代谢转化成全反式维甲酸（ATRA），ATRA 通过一系列的氧化及异构化作用进一步代谢生成多种有活性的类维生素 A，进而行使其细胞内的生理功能[30]。

在正常培养条件下的小鼠 ES 细胞，极少有外源维生素 A 发生代谢。一旦撤除 LIF 因子，ES 细胞发生外源性代谢变化，维生素 A 转化成 4-羟基视黄醛（4-hydroxyretinol）和 4-氧化视黄醛（4-oxoretinol），同时 ES 细胞中分化标志基因 FGF-5（fibroblast growth factor-5）的表达被激活，而干细胞标志基因 FGF-4 的表达降低，ES 细胞开始分化。这表明在撤除 LIF 的环境下，ES 细胞的分化与视黄醛的代谢紧密相关[31]。

RA 是视黄醛的生理代谢产物。在脊椎动物胚胎发育过程中，原肠期时后脑的发育对过量的 RA 敏感[25]，RA 过量，颅面及中枢神经系统、肢体的发育会产生畸变[32~34]。早期心脏发育对类维生素 A 的浓度敏感，在心肌细胞分化中 RA

对肌节蛋白表达模式的重组起重要作用[29]。RA 影响分化及相关组织的细胞外基质（extracellar matrix）的代谢[35]。RA 通过特异性结合细胞内蛋白并与核内受体作用介导特定基因的转录，进而直接影响靶基因的表达，调节细胞分化[36]。RA 诱导 ES 细胞分化可以产生大量的神经细胞，同时也分化形成其他类型细胞。研究中发现在胚胎内分化区域表达的基因在 RA 诱导的 EB 分化中也有表达，这说明 RA 在体外的诱导作用比仅诱导神经基因的表达更广泛。但无论是体内还是体外，RA 如何诱导神经发生的机制都不清楚。近年来，许多研究者兴趣集中于用多能干细胞来源的神经元为实验模型体外研究神经系统分化的早期分子事件[37]。选择标记与特定启动子相连，而此启动子仅在感兴趣的细胞群中有活性。这样在分化过程中，表达标记基因的细胞可以通过细胞分类或药物筛选得以纯化[38,39]。

国内外对神经系统的分化及发生的研究主要集中于一种或少数几种已知基因及因子对此途径的影响，并侧重于神经干细胞的终末分化。任何一种生物，特别是高等生物生命过程的分子基础是相当复杂的，研究了解这一过程的最好方法是分析不同发育阶段不同组织、正常与病理条件下基因的差异表达。利用传统的基因分析方法，如基因剔除等，操作过程复杂且一次只能分析一个基因。基于上述原因，近年来发展起来的蛋白质组学便受到了普遍的关注。从蛋白质组研究的角度研究 ES 细胞体外定向分化的细胞模型，可对处于同一分化阶段的调控因子同时做分析比较，有利于全面研究胚胎发生较早期神经发生与分化过程中的分子事件。由于早期神经发生在胚胎器官发生中的重要的位置效应及诱导效应，人们致力于研究神经发生的早期基因调控，以及参与细胞早期决定与分化的分子机制。尽管现在人们对于神经系统的发育与分化已有一定的了解，但对于胚胎发生的早期事件及基因的机制仍不十分清楚。如果能够建立由 ES 细胞向神经细胞分化的体外模型，那将对研究神经发生有极大的帮助。首先，可以在神经元发育的关键阶段点分离得到大量的细胞以供研究；其次，可以在组织培养条件下模拟体内发育关键步骤，系统地改变和控制影响神经元发育的外源因子，利用此模型找出神经系统发生的诱导因子及介导者，确定在神经元及神经胶质细胞分化中基因的功能；再次，可以建立神经退行性病变的模型，利用 ES 细胞在特定位点改变基因，通过对特定基因的剔除、改变、增加来模拟普通遗传疾病的病理发生，发现治疗神经性病变药物及研究细胞移植治疗和基因治疗等等。另外，关于神经细胞的一个令人注意的现象是最近发现的神经细胞与其他类型细胞（主要是造血细胞系）之间跳跃式的命运互换[40,41]，这使得人们对于传统的细胞命运决定、分化及去分化的观念需要做重新认识。

五、小鼠 ES 细胞向神经细胞分化的比较蛋白质组研究

目前，一些高等生物的基因组研究取得了日新月异的进展，但人类对这些遗

传密码基因的功能所知还甚少，于是功能基因组的研究成为新的趋势。蛋白质组（proteome）的概念是 Marc Wilkins[42]于 1994 年率先提出的，其定义为“一种基因组所表达的全套蛋白质”，由其衍生的 protcomics，即蛋白质组学，意指大规模研究蛋白质的表达水平变化、翻译后修饰和蛋白质与蛋白质相互作用，它是对基因组学的重要补充，是在蛋白质水平定量、动态、整体性研究生物体，以获取生物疾病病变、细胞进程及蛋白质网络的整体综合信息的科学研究。与以往蛋白质化学的研究不同，蛋白质组研究的对象不是单一或少数的蛋白质，它侧重的是全面性和整体性，需要获得体系内所有蛋白质组分的物理、化学及生物学参数，如相对分子质量、等电点、表达量等。蛋白质是细胞的功能单位，蛋白质表达反映了细胞目前的状态。蛋白质研究可能给出与基因组研究相似的结果，也可以克服蛋白表达与 mRNA 水平的差异[43]，检测出翻译后修饰如糖基化、磷酸化，蛋白分子断裂及其他多种翻译后修饰[44]。

当前蛋白质组分析的核心技术是双向聚丙烯酰胺凝胶电泳（two-dimensional electrophoresis，2-DE）。2-DE 用于分离复杂的蛋白质组分，目前多应用固相 pH 梯度电泳等电聚焦为第一向分离蛋白质，然后用 SDS-PAGE 按相对分子质量分离蛋白质。通过银染和放射自显影显色，经扫描后利用专门的计算机软件对所得的图像进行数据采集和分析，可获得蛋白质等电点、相对分子质量、表达量等方面的信息，进行不同环境下蛋白表达模式的比较。互连网现在提供了许多参照胶（如 SWISS-2DPAGE，Ref52 database）[45,46]，使不同类型组织及不同生长状况下的生物体可以直接进行蛋白质比较。利用氨基酸组成分析、微量蛋白质序列分析、质谱技术等对从胶上回收的蛋白质斑点进行精细的鉴定，可以获得有关蛋白质性质、表达变化及翻译后加工等大规模信息[47~49]。

用肽质量指纹谱（peptide mass fingerprinting，PMF）鉴定蛋白质是目前蛋白质组研究较为常用的鉴定方法，肽质量指纹谱是指蛋白质被识别特异酶切位点的蛋白酶水解后得到的肽片段质量图谱。由于每种蛋白的氨基酸序列（一级结构）都不同，当蛋白被酶水解后，产生的肽片段序列也各不相同，其肽混合物质量数亦具特征性，称为肽质量指纹谱（PMF）[50~53]。肽质量指纹谱鉴定技术主要有三个步骤：① 胶上蛋白原位酶切得到肽混合物；② 肽混合物质谱分析得到肽质量指纹谱；③ 用肽指纹谱在蛋白数据库中检索蛋白质。随着质谱技术的不断发展，对生物大分子的分析可以更微量和更精确，新一代的质谱仪器分析肽混合物灵敏度可达 fmol，精确度可达万分之一，从而从技术上推进了蛋白质组的研究。

通过 2-DE 分析蛋白质组的方法仍不断地改善，现在生物学家的兴趣集中于研究同时表达、调控和翻译后修饰的成百上千个蛋白。而进一步完善的 2-DE 分离及分离后的分析技术无疑将促使蛋白质组研究更快更方便，这些分离和分析技术与蛋白质数据库、基因数据库的结合将进一步推动基因组和蛋白质组的研究和

发展。综合起来，蛋白质组研究的技术体系包括对细胞内所有蛋白质的分离，蛋白表达模式的识别，蛋白的鉴别和认定，及蛋白的翻译后修饰分析。

为了检测 RA 诱导神经细胞分化的分子机制，并进一步研究神经分化早期的分子调控机制，我们分别收集了 RA 作用不同时期的分化细胞的总蛋白质，应用 2-DE 技术分离、分析，建立了 ES 细胞和不同分化时期神经细胞的蛋白质图谱。处于同一分化阶段的调控因子同时做分析比较，有利于全面研究胚胎发生较早期神经发生与分化和其他细胞组织系统发生之间潜在的联系，将改变以往基因操作的复杂及繁琐，并能弥补基因操作技术只能同时研究一个或少数几个基因功能的不足，以期了解胚胎发生中向神经系统分化过程早期的分子事件[54]。

1. ATRA 体外高效诱导 ES 细胞向神经细胞分化

在体外 ES 细胞自发分化成为神经细胞是很少见的。为了增加从 ES 细胞衍生出的神经元和胶质细胞的数量，我们对无饲养层细胞和 LIF 培养基培养的 ES 细胞设计了一个分化步骤（图 12-2）。

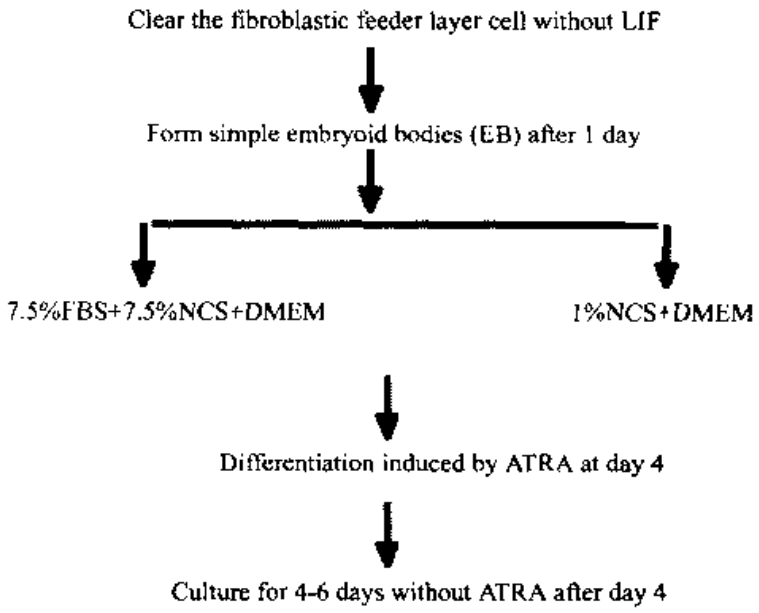


图 12-2　简单分化步骤

首先，ES 细胞的培养：用丝裂霉素 C 处理的小鼠胚胎成纤维细胞（mouse embryonic fibroblast，MEF）作为饲养细胞层，用 ES 细胞正常培养液培养。每天换液，隔天传代。根据细胞的生长情况及实验要求的不同按不同比例传代。其次，胚胎成纤维细胞的制备：将孕期 13.5 + / − 天的昆明小鼠引颈处死，70% 酒精消毒，无菌操作。取出胚胎，剥去胚胎外膜，去除胎肝、脑、四肢等。将胚胎剪碎，4 号针头过滤，用 PBS 冲洗 3 次，小心吸弃上层破碎细胞。将细胞转移至新的培养瓶中，加 1ml 0.025% 胰蛋白酶 − 0.04% EDTA，反复吹打，混匀，转移

上清液至离心管中。1000r/min 离心 5min，弃上清，加入培养液使细胞重悬浮，铺皿培养，记为 0 代 MEF 细胞，37℃孵育。第三，饲养层细胞的制备：将处于增长期的 1～4 代 MEF 细胞的培养液吸弃，加入 3ml 含 10% 小牛血清和 10μg/ml 丝裂霉素 C 的 DMEM，37℃孵育 2h。第四，BRL 条件培养液的制备：BRL 细胞以 1×10^5/ml 密度接种，用含 7.5% NCS 的 DMEM 培养，72h 后收集培养液，3000r/min 离心 20min，去细胞沉淀，每三天收集一次，收集 3 周。离心后的培养液经过 0.22μm 的微孔滤膜过滤，加 7.5% 的胎牛血清，作为 A 液，常规培养 ES 细胞的培养液为 B 液，混合两者比例（A/B）为 60/40（V/V），再加 β-巯基乙醇至终浓度为 10^{-4}mol/L。第五，饲养细胞的去除：胰酶消化含饲养细胞的胚胎干细胞，换用 BRL 条件培养液培养，静置 4h 后，小心地将未贴壁的胚胎干细胞转移至 0.1% 明胶处理的培养瓶培养。最后，定向诱导 ES 细胞分化：在不黏附底物之中悬浮培养 4d，形成简单的胚胎小体（embryoid bodies，EBs），EBs 在含有 5×10^{-7}mol/L ATRA 的低浓度（1%）胎牛血清中培养，直至 4d，已分化的细胞转至不含 LIF 和 FBS 的正常培养基中，培养至少 10d。随后，我们利用免疫组化的方法检测抗神经纤维抗体（仅在成熟神经细胞中表达）。结果显示，神经纤维细胞的阳性率超过 80%（图 12-3）。因此表明体外定向诱导 ES 向神经细胞分化的模型已经建立起来。

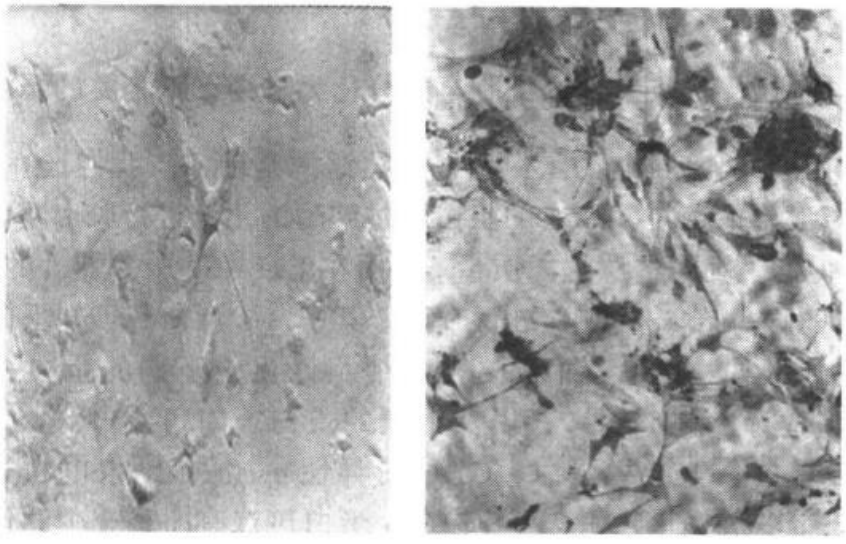


图 12-3　免疫组化检测 ES 细胞分化为神经细胞
左为对照组；右为分化组

2．利用考马斯亮蓝染色的双向电泳分离与 ES 细胞早期分化为神经细胞的相关蛋白

为了鉴定 ES 细胞早期分化为神经细胞的相关蛋白，通过双向电泳技术，对

未分化的 ES 细胞和 ATRA 诱导的已分化的神经细胞的蛋白表达谱进行比较。选择 24 个差异表达的蛋白斑点，其中，蛋白斑点 No.1～10 仅在 ES 细胞中表达，但在 RA4 细胞中没有或微量表达；与之相反，蛋白斑点 No.11～24 仅在 RA4 细胞中表达，但在 ES 细胞中没有或低表达（图 12-4）。显然，这些蛋白与 ES 细胞早期分化为神经细胞的蛋白相关。需要进一步通过质谱进行鉴定。

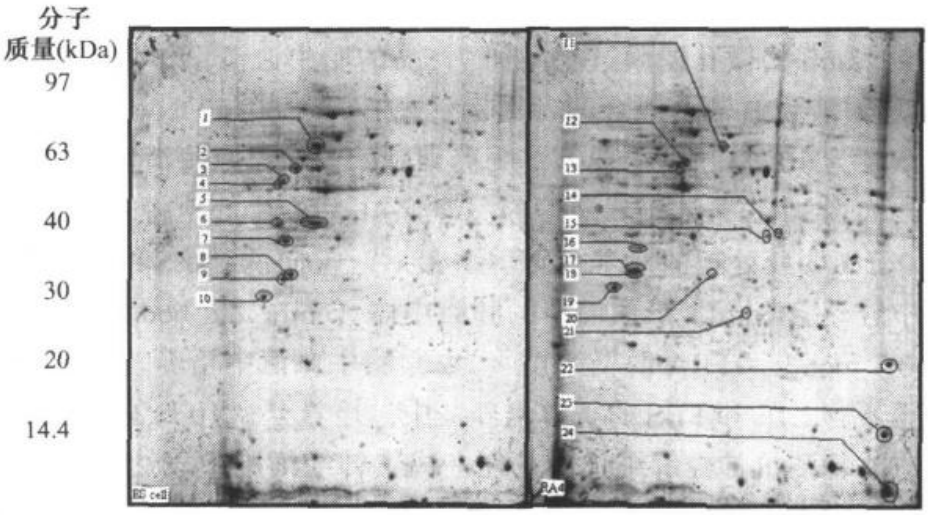


图 12-4　双向电泳分离与 ES 细胞早期分化为神经细胞的相关蛋白

3．通过肽质量指纹谱鉴定与 ES 细胞早期分化为神经细胞相关的蛋白

在比较 ES 和 RA4 细胞的蛋白质组表达谱的基础上，我们选择了 24 个蛋白进行肽质量指纹谱的鉴定。所有的蛋白经过胰酶消化，进行胶上酶解，并且通过 PMF 分析获得近乎完美的肽片段质量。利用在 ExPASy 服务器上的 PeptIdent 软件（http：//www.expasy.ch/）进行查询，其中 9 个蛋白与已知蛋白具有高配比率（表 12-1），例如：在 ES 细胞向神经细胞分化中下调的蛋白——波形蛋白（vimentin）和微管蛋白 α-3/α-7；以及上调的蛋白，如角蛋白 8、18、血小板激活因子 IBα 亚单位、NEL 样蛋白、G1/S 特异性 cyclin D2、follistatin-related protein 和硫氧还蛋白过氧化物酶 2；另有 3 个蛋白斑点有超过一个的候选者，不能明确。此外，剩下的 12 个蛋白，如在 ES 细胞向神经细胞分化中 5 个上调和 7 个下调的蛋白，未能从数据库中得到结果，提示可能为未知蛋白。

通过对不同分化时期细胞蛋白质组的比较，我们随机选取了一些蛋白差异点做肽质量指纹谱分析，初步鉴定了一些与神经分化相关的蛋白质，包括细胞骨架蛋白、follistatin 相关蛋白、类 NEL 蛋白、血小板激活因子乙醇水解酶和硫氧还蛋白过氧化氢酶等（见表 12-1）。

表 12-1 **Results of PeptIdent search of identified protein spots**

Spot No.	Match rate	Description	pI	Mw (kDa)	Relative quantity (G_{ES}/G_{RA4})
1	14/18	Vimentin	5.06	53.5	10/1
7	6/11	Tubulin A-3/A-7	4.98	49.9	6/2
11	9/11	Cytokeratin 8	5.50	54.3	0/4
12	9/13	Cytokeratin 18	5.22	47.3	0/6
14	8/14	Platelet-activating factor acetylhydrolase IB alpha subunit	7.03	46.5	0/2
16	8/11	NEL-like protein (fragment)	Undefined	Undefined	1/4
18	6/11	G1/S-specific cyclin D2	5.04	32.8	0/3
20	6/11	Follistatin-related protein	5.41	32.4	0/1
22	8/12	Thioredoxin peroxidase 2	8.26	22.2	0/10
	7/15	Amiloride-sensitive sodium channel alpha-subunit	Undefined	Undefined	
9		(fragment)			2/1
	6/15	Diff6 protein	5.70	41.9	
	6/15	H-2 class I histocompatibility antigen, K-Q alpha chain (fragment)	Undefined	Undefined	
	5/12	Chain 1: 30 kD adipocyte	5.19	24.9	
19		complement-related			0/4
	5/12	Sorbitol dehydrogenase	Undefined	Undefined	
23	4/11	Cofilin, non-muscle isoform	8.22	18.6	0/10
	4/11	Ig heavy chain V region HPCG8	Undefined	Undefined	

脊椎动物神经系统早期发育的特征是形态发生运动过程和细胞分化过程，这两者奠定了神经发育的基础。在神经分化过程中有明显的细胞内骨架蛋白表达的变化，如中间纤维蛋白[55]。神经胚形成早期，神经管细胞一般表达波形纤维蛋白[55~57]。波形纤维蛋白是一种在多种组织中表达的中间纤维蛋白，它与 nestin 和 RC1 抗原被认为是神经干细胞的标记分子[58~60]。波形纤维蛋白的出现标志着神经分化的开始。对鼠胚早期发育所进行的形态、免疫、电泳等研究表明，在胚胎发育的第 11 天，神经外胚层中，角蛋白迅速减少或停止表达，同时出现波形纤维蛋白。一些将要发育为星形细胞的在 18 天开始同时表达波形纤维蛋白和胶质纤维蛋白；而要发育为神经细胞的则在角蛋白向波形纤维蛋白转变后很短时间

内即开始表达神经纤维蛋白。神经纤维蛋白与波形纤维蛋白共同表达 2～3 天，继而波形纤维蛋白表达下降和停止[61,62]。我们利用蛋白质组研究技术初步从蛋白水平上证实了波形纤维蛋白在神经分化过程中表达量的变化，在未分化的 ES 细胞中波形纤维蛋白少量表达。加入 RA，启动神经分化过程，波形纤维蛋白表达量增高；而到了分化末期的 RA-4 细胞中，波形纤维蛋白的表达降低，与以往的研究结果相符[62]。

微管蛋白一般在发育的神经组织中表达，但在我们的分析鉴定中，发现 ES 细胞中亦有微管蛋白 α-3/α-7 链的表达，而在 RA4 细胞中没有检测到其表达。这可能是 ES 来源的神经细胞一般表达 β 型微管蛋白[37]。

现在有许多报道指出一些生长因子可以影响神经细胞的增殖和分化[63,64]。活化素是一种有丝分裂原，是 TGFβ 家族中的一员，它可以影响多种类型的体外培养细胞的生长和分化。在胚胎发育中，内源性活化素参与中胚层的诱导、体轴形成及器官发生[65,66]。作为神经发生的抑制因子[67,68]，活化素可以在体外抑制 RA 诱导的 P19 畸胎瘤细胞系及一些神经细胞瘤向神经细胞分化[69]。Follistatin 是一种活化素结合蛋白，可以通过结合活化素抑制其活性，参与神经系统分化的过程。体外培养的 P19 细胞用 RA 处理后，会出现暂时的 follistatin 表达高峰，48h 后 mRNA 的表达消失，在分化后期（120h 后），又有较明显的 follistatin mRNA 的表达[70]。由于 follistatin 对 RA 反映非常快速，说明 follistatin 基因可能是一种 RA 调节的基因。Follistatin 自身也可以刺激神经细胞的分化。神经细胞的增殖和分化受活化素和 follistatin 的控制，活化素和 follistatin 的平衡关系可能是对维持 P19 细胞处于一种增生状态所必需的。我们在 RA-4 中鉴定的 No.20 蛋白质 follistatin 相关蛋白在体外神经分化过程中具有与 follistatin 在 RA 诱导 P19X 细胞中相似的表达模式，说明 follistatin 相关蛋白和 follistatin 一样，可能是一种 RA 调节的蛋白。RA 处理 P19 细胞后 follistatin 的 mRNA 表现出迅速的短暂性增强，虽然此过程的生理意义还不十分清楚，但 follistatin 作为肝磷脂结合蛋白，可能通过与细胞表面的蛋白多糖结合来调节活化素的活性，从而参与对神经分化过程的调节[71]。

类 NEL 蛋白也是一种与神经元分化相关的蛋白质，它与已知的 NEL 蛋白结构类似，都含有类 EGF 重复片段[72]。在我们的体外模型系统中，类 NEL 蛋白在 ES 细胞中没有表达，而在 RA-4 细胞中表达。这和最早发现的鸡胚中 NEL 的表达不很一致。在鸡胚中，多种组织中都有 NEL 表达，但孵化后，则只在大脑和视网膜表达。这既可能是体内外差异的结果，也可能是不同种属间同源蛋白时空表达模式的不同所表现的不同。现在 NEL 的功能不十分清楚，一般认为作为神经元特殊信号传导途径的一个受体或配体行使其功能[73]。我们认为类 NEL 蛋白可能作为一种配体刺激其他分子如 EGF 受体来行使其在神经分化中的作用。

血小板激活因子（PAF）是一种有效的生物内分泌素，参与多种生理过

程[74,75]。在神经组织及神经来源的细胞，PAF 调节神经递质的释放[76,77]与突触的传接[78,79]。PAF 是一种重要的神经调节子，通过 G 蛋白特定受体的介导在多种组织内参与信号传导途径。PAF 乙醇水解酶能抑制 PAF 的活性。PAF 乙醇水解酶与 PAF 协调作用，共同参与神经分化及形成。在我们的实验结果中，PAF 乙醇水解酶在 ES 细胞中微量表达，而在 RA-4 时期的细胞中表达量有所增高。从而表明与其已知的功能相一致。

我们还鉴定了一种与细胞周期调控有关的蛋白，即 G1/S 特异性蛋白 cyclin D2，它特异性地调控细胞分裂由 G1 期进入 S 期，这个位点是哺乳动物细胞分裂的限制位点，进入 S 期后，细胞开始 DNA 的复制，为细胞分裂做准备。在我们的体外分化系统中，RA-4 期细胞中仍有大量的 cyclin D2 表达，说明这个时期的分化细胞并没达到分化终点，形成功能性的神经元，而可能是处于神经干细胞进一步分化产生定向干细胞的时期。ES 细胞的分裂也很旺盛，但并未检测到有这种 cyclin D2 的表达，这可能是不同类型的 cyclin D2 在不同的组织中表达的缘故[80,81]。

综上所述，我们成功地建立了由 RA 诱导的 ES 细胞体外定向向神经分化的模型系统。并利用 2-DE 技术分离、分析了 ES 细胞向神经系统分化过程几个时相点的细胞的总蛋白质，比较其中的差异蛋白点，并对其中几个蛋白点作了肽质量指纹谱鉴定。与已知肽库的蛋白质做比较，初步确定了几种与神经细胞分化及细胞增殖相关的蛋白质，并且发现它们在 ES 细胞向神经系统分化过程中的时空表达模式与已有的体内外研究结论相符。这一方面证明了我们所建模型在生物学上的可靠性，另一方面也证明了从蛋白质组角度来分析研究体外分化模型的可行性，上述研究模型与分析体系将有力推动体内胚胎早期发生与分化分子机制的深入研究。将胚胎干细胞来源的体外分化模型系统与蛋白质组研究技术相结合，不但可以克服体内研究的诸多困难，还能利用蛋白质组研究的优点对多个基因及因子同时做比较分析，这有效地提高工作效率，大大减少以往基因操作中人力、物力的消耗及因研究对象的单一性所带来的结果的固有的局限性。本实验还鉴定了一些蛋白，它们不能与现有肽库中的蛋白相匹配，鉴于上述系统的有效性，提示很可能是未知蛋白，如果通过氨基酸测序、cDNA 文库筛选及 RT-PCR 等手段克隆其 cDNA，并进一步研究它们在神经分化过程中的作用，无疑将可能为神经系统发育与分化相关的分子机制提供新的切入点或更深入系统的理论认识。

参 考 文 献

1 Odorico J S, Kaufman D S, Thomson J A. Multilineage differentiation from human embryonic stem cell lines. Stem Cells, 2001, 19 (3): 193－204

2 Watt F M, Hogan B L. Out of Eden: stem cells and their niches. Science, 2000, 287: 1427－1430

3 Evans M J, Kaufman M H. Establishment in culture of pluripotential cells from mouse embryos. Nature, 1981, 292 (9): 154－156

4 Gall R Martin. *Isolation of a pluripotential Cell line from early mouse embryos cultured in medium conditioned* by teratocarcinoma stem cells. Proc Natl Acad Sci USA, 1981, 78 (12): 7634－7638

5 Prelle K, Vassiliev I M, Vassilieva S G, Wolf E, Wobus A M. Establishment of pluripotent cell lines from vertebrate species: present status and future prospects. Cell Tissues Organs, 1999, 165: 220－236

6 Rohwedel J, Guan K, Hegert C, Wobus A M. Embryonic stem cells as an in vitro model for mutagenicity, cytotoxicity and embryo toxicity studies: present state and future prospects. Toxicol In Vitro, 2001, 15 (6): 741－753

7 Gokhan S, Mehler M F. Basic and clinical neuroscience applications of embryonic stem cells. Anat Rec, 2001, 265 (3): 142－156

8 Pesce M, Scholer H R. Oct-4: gatekeeper in the beginnings of mammalian development. Stem Cells, 2001, 19 (4): 271－278

9 Niwa H, Miyazaki J, Smith A G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet, 2000, 24 (4): 372－376

10 Keller G M. In vitro differentiation of embryonic stem cells. Curr Opin Cell Biol, 1995, 7: 862－869

11 Rohwedel J, Maltsev V, Bober E, Arnold H H, Hescheler J, Wobus A M. Muscle cell differentiation of embryonic stem cells reflects myogenesis in vivo: developmentally regulated expression of myogenic determination genes and functional expression of ionic currents. Dev Biol, 1994, 164: 87－101

12 Leahy A, Xiong J W, Kuhnert F, Stuhlmann H. Use of developmental marker genes to define temporal and spatial patterns of differentiation during embryoid body formation. J Exp Zool, 1999, 284: 67－81

13 Rathjen P D, Lake J, Whyatt L M, Bettess M D, Rathjen J. Properties and uses of embryonic stem cells: prospects for application to human biology and gene therapy. Reprod Fertil Dev, 1998, 10: 31－47

14 Wobus A M, Grosse R, Schoneich J. Specific effects of nerve growth factor on the differentiation pattern of mouse embryonic stem cells in vitro. Biomed Biochim Acta, 1988, 47: 965－973

15 Bain G, Ray W J, Yao M. Retinoic acid promotes neural and represses mesodermal gene expression in mouse embryonic stem cells in culture. Biochem Biophys Res Commun, 1996, 223: 691－694

16 Fraichard A, Chassande O, Bilbaut G, Dehay C, Savatier P, Samarut J. In vitro differentiation of embryonic stem cells into glial cells and functional neurons. J cell science, 1995, 108: 3181－3188

17 Okabe S, Forsberg-Nilsson K, Spiro A C, Segal M, McKay R D. Developmental of neuronal precursor cells and functional postmitotic neuron from embryonic stem cells in vitro. Mech Dev, 1996, 59: 89－102

18 Shamblott M J, Axelman J, Wang S, Bugg E M, Littlefield J W, Donovan P J, Blumenthal P D, Huggins G R, Gearhart J D. Derivation of pluripotent stem cells from cultured human primordial germ cells. Proc Natl Acad Sci USA, 1998, 95 (23): 13726－13731

19 Matsui Y, Zsebo K, Hogan B L. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell, 1992, 70 (5): 841－847

20 Amit M, Carpenter M K, Inokuma M S, Chiu C P, Harris C P, Waknitz M A, Itskovitz-Eldor J, Thomson J A. Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. Dev Biol, 2000, 227 (2): 271－278

21 Thomson J A, Marshall V S. Primate embryonic stem cells. Curr Top Dev Biol, 1998, 38: 133－165

22 Thomson J A, Itskovitz-Eldor J, Shapiro S S, Waknitz M A, Swiergiel J J, Marshall V S, Jones J M. Embryonic stem cell lines derived from human blastocysts. Science, 1998, 282 (6): 1145－1147

23 Reubinoff B E, Pera M F, Fong C Y, Trounson A, Bongso A. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. Nat Biotechol, 2000, 18: 399－404

24 Itskovitz-Eldor J, Schuldiner M, Karsenti D, Eden A, Yanuka O, Amit M, Soreq H, Benvenisty N. Differ-

entiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. Mol Med, 2000, 6 (2): 88～95

25 Schuldiner M, Yanuka O, Itskovitz-Eldor J, Melton D A, Benvenisty N. From the cover: effects of eight growth factors on the differentiation of cells derived from human embryonic stem cells. Proc Natl Acad Sci USA, 2000, 97 (21): 11307～11312

26 Keaveney M, Stunnenberg H G. Retinoic acid receptors. In: Baeurle PA, ed. Inducible Gene Expression. Vol 2. Boston: Birkhauser, 1995. 197～242

27 Wilson J G, Roth C B, Warkany J. An analysis of the syndrome of malformations induced by maternal vitamin A deficiency: effects of restoration of vitamin A at various times during gestation. Am J Anat, 1953, 92: 189～217

28 Goulding E H, Pratt R M. Teratogenicity in whole embryo culture. J Craniofac Genet Dev Biol, 1986, 6: 99～112

29 Mitrani E, Shimoni Y. Retinoic acid inhibits growth in agarose of early chick embryonic cells and may be involved inregulation of axis formation. Development, 1989, 107: 275～280

30 Warrell R P Jr, de The H, Wang Z Y, Degos L. Acute promyelocytic leukemia. N Engl J Med, 1993, 329: 117～189

31 Lane M A, Chen A C, Roman S D, Derguini F, Gudas L J. Removal of LIF (leukemia inhibitory factor) results in increased vitamin A (retinol) metabolism to 4-oxoretinol in embryonic stem cells. Proc Natl Acad Sci USA, 1999, 96 (23): 13524～13529

32 Shenefelt R E. Morphogenesis of malformations in hamsters caused by retinoic acid: relation to dose and stage at treatment. Teratology, 1972, 5: 103～118

33 Maden M, Gale E, Kostetskii I, Zile M. Vitamin A-deficient quail embryos have half a hindbrain and other neural defects. Curr Biol, 1996, 6: 417～426

34 Sharpe C R, Goldstone K. Retinoid receptors promote primary neurogenesis in Xenopus. Development, 1997, 124: 515～523

35 Oikarinen A, Vuorio T, Makela J, Vuorio E. 13-cis retinoic acid and dexamethasone modulate the gene expression of epidermal growth factor and fibroblast proteoglycan 40 core protein in human skin fibroblast. Acta Derm Venereol (Stockh), 1989, 69: 466～469

36 Doevendans P A, Bilsen van M. Transcription factors and the cardiac gene programme. Int J Biochem Cell Biol, 1996, 28: 387～403

37 Bain G, Kitchens D, Yao M, Huettner J E, Gottlieb D I. Embryonic stem cells express neuronal properties in vitro. Developmental Biology, 1995, 168: 342～357

38 Li M, Pevny L, Lovell-Badge R, Smith A. Generation of purified neural precursors from embryonic stem cells by lineage selection. Curr Biol, 1998, 8 (17): 971～974

39 Klug M G, Soonpaa M H, Koh G Y, Field L J. Genetically selected cardiomyocytes from differentiating embryonic stem cells form stable intracardiac grafts. J Clin Invest, 1996, 98: 216～224

40 Bjornson C R, Rietze R L, Reynolds B A, Magli M C, Vescovi A L. Turning brain into blood: a hematopoietic fate adopted by adult neural stem cells in vivo. Science, 1999, 283 (5401): 534～537

41 Azizi S A, Stokes D, Augelli B J, DiGirolamo C, Prockop D J. Engraftment and migration of human bone marrow stromal cells implanted in the brains of albino rats——similarities to astrocyte grafts. Proc Natl Acad Sci USA, 1998, 95 (7): 3908～3913

42 Wasinger V C, Cordwell S J, Cerpa～Poljak A, Yan J X, Gooley A A, Wilkins M R, Duncan M W, Harris R, Williams K L, Humphery-Smith I. Progress with gene-product mapping of the Mollicutes: Mycoplasma

genitalium. Electrophoresis, 1995, 16: 1090～1094

43 Anderson L, Seilhamer J A. Comparison of selected mRNA and protein abundances in human liver. Electrophoresis, 1997, 18: 533～537

44 Humphery-Smith I, Blackstock W. Proteome analysis: genomics via the output rather than the input code. J Protein Chem, 1997, 16 (5): 537～544

45 Sanchez J C, Appel R D, Golaz O, Pasquali C, Ravier F, Bairoch A, Hochstrasser D F. Inside SWISS-2DPAGE database. Electrophoresis, 1995, 16: 1131～1151

46 Boutell T, Garrels J I, Franza B R, Monardo P J, Latter G I, Boutell T, Garrels J I, Franza B R, et al. REF52 on global gel navigator: An internet-accessible two-dimensional gel electrophoresis database. Electrophoresis, 1994, 15: 1487～1490

47 Matsudaira P. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J Biol Chem, 1987, 262 (21): 10035～10038

48 Mann M, Wilm M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Anal Chem, 1994, 66: 4390～4399

49 Gold M R, Yungwirth T, Sutherland C L, Ingham R J, Vianzon D, Chiu R, van Oostveen I, Morrison H D, Aebersold R. Purification and identification of tyrosine phosphorylated proteins from B lymphocytes stimulated through the antigen receptor. Electrophoresis, 1994, 15: 441～453 .

50 Humphery-Smith I, Cordwell S J, Blackstock W P. Proteome research: Complementarity and limitations with respect to the RNA and DNA worlds. Electrophoresis, 1997, 18: 1217～1242

51 Yates J R 3rd. Mass spectrometry and the age of the proteome. J Mass Spectrom, 1998, 33 (1): 1～19

52 Jungblut P, Thiede B. Protein identification from 2-DE gels by MALDI mass spectrometry. Mass Spectrom Rev, 1997, 16 (3): 145～162

53 Dainese P, Staudenmann W, Quadroni M, Korostensky C, Gonnet G, Kertesz M, James P. Probing protein function using acombination of gene knockout and proteome analysis by masss pectrometry. Electrophoresis, 1997, 18: 432～442

54 Guo X, Ying W, Wan J, Hu Z, Qian X, Zhang H, He F. Proteomic characterization of early-stage differentiation of mouse embryonic stem cells into neural cells induced by all-trans retinoic acid in vitro. Electrophoresis, 2001, 22 (14): 3067～3075

55 Tapscott S J, Bennett G S, Toyama Y, Kleinhart F, Holtzer H. Intermediate filament proteins in the developing chick spinal cord. Dev Biol, 1981, 86: 40～54

56 Houle J, Fedoroff S. Temporal relationship between the appearance of vimentin and neural tube development. Dev Brain Res, 1983, 9: 189～195

57 Stagaard M, Moll～rd K. The developing neuroepithelium in human embryonic and fetal brain studied with vimentin-immunocytochemistry. Anat Embryol, 1989, 180: 17～28

58 Frederiksen K, Mckay R D G. Proliferation and differentiation of rat neuroepithelial of bipotential and plastic glio-neuronalprecursor cells. J Neurosci, 1988, 8: 1144～1151

59 Lendahl U, Zimmerman L B, Mckay R D G. CNS stem cells express a new class of intermediate filament protein. Cell, 1990, 60: 585～595

60 Edwards M A, Yamamoto K, Bishop J M. Organization of radial glia and related cells in the developing murine CNS. An analysis based upon a new monoclonal antibody marker. Neuroscience, 1990, 36: 121～144

61 Bennett G S, DiLullo C. Expression of a neurofilament protein by the precursors of a subpopulation of ventral spinal cord neurons. Dev Biol, 1985, 107: 94～106

62 Cochard P, Paulin D. Initial expression of neurofilaments and vimentin in the central nervous system of the mouse embryo in vivo. Neurosci, 1984, 4: 2080～2094

63 Cattaneo E, Mckay R. Proliferation and differentiation of neuronal stem cells regulated by nerve growth factor. Nature, 1990, 347: 762～765

64 McKinnon R D, Matsui T, Dubois-Dalcq M, Aaronson S A. FGF modulates the PDGF-driven pathway of oligo dendrocyte development. Neuron, 1990, 5: 603～614

65 Mitrani E, Ziv T, Thomsen G, Shimoni Y, Melton D A, Bril A. Activins are expressed early in Xenopus embryogenesis and can induce axial mesoderm and anterior structures. Cell, 1990, 63: 485～493

66 Ruiz I, Altaba A, Melton D A. Axial patterning and the establishment of polarity in frog embryo. Trends Genet, 1990, 6: 57～64

67 Hashimoto M, Kondo S, Sakurai T, Etoh Y, Shibai H, Muramatsu M. Activin/EDF as an inhibitor of neural differentiation. Biochem Biophys Res Commun, 1990, 173: 193～200

68 Hemmati-Brivanlou A, Melton D A. Inhibition of activin receptor signaling promotes neuralization in Xenopus. Cell, 1994, 77: 273～281

69 Van den Eijnden-van Raaij A J, van Achterberg T A, van der Kruijssen C M, Piersma A H, Huylebroeck D, de Laat S W, Mummery C L. Differentiation of aggregated murine P19 embryonal carcinoma cells is induced by a novel visceral endoderm-specific FGF-like factor and inhibited by activin A. Mech Dev, 1991, 33 (2): 157～165

70 Hashimoto M, Nakamura T, Inoue S, Kondo T, Yamada R, Eto Y, Sugino H, Muramatsu M. Follistatin is a developmentally regulated cytokine in neural differentiation. J Biol Chem, 1992, 267 (11): 7203～7206

71 Nakamura T, Sugino K, Titani K, Sugino H. Follistatin, an activin-binding protein, associates with heparan sulfate chains of proteoglycans on follicular granulosa cells. J Biol Chem, 1991, 266 (29): 19432～19437

72 Watanabe T K, Katagiri T, Suzuki M, Shimizu F, Fujiwara T, Kanemoto N, Nakamura Y, Hirai Y, Maekawa H, Takahashi E. Cloning and characterization of two novel human cDNAs (NELL1 and NELL2) encoding proteins with six EGF-like repeats. Genomics, 1996, 38: 273～276

73 Matsuhashi S, Noji S, Koyama E, Myokai F, Ohuchi H, Taniguchi S, Hori K. New gene, nel, enconding a Mr 93K protein with EGF-like repeats is strongly expressed in neural tissues of early stage chick embryo. Dev Dyn, 1995, 203: 212～222

74 Kornecki E, Ehrlich Y H. Neuroregulatory and neuropathological activating actions of the ether-phospholipid platelet-activating factor. Science, 1998, 240: 1792～1794

75 Panetta T, Marcheselli V L, Braquet P, Spinnewyn B, Bazan N G. Effects of a platelet-activating factor antagonist (BN52021) on free fatty acids, diacylglycerols, polyphosphoinositides and blood flow in the gerbil brain: inhibition of ischemia-reperfusion induced cerebral injury. Biochem Biophys Res Commun, 1987, 149: 580～587

76 Bussolino F, Tessari F, Turrini F, Braquet P, Camussi G, Prosdocimi M, Bosia A. Platelet-activating factor induces dopamine release in PC-12 cell line. Am J Physiol, 1988, 255: C559～C565

77 Wang H Y, Yue T L, Feuerstein G, Friedman E. Platelet-activating factor: diminished acetylcholine release from rat brain slices is mediated by a Gi protein. J Neurochem, 1994, 63: 1720～1725

78 Cerro S D, Arai A, Lynch G. Inhibition of long term potentiation by an antagonist of platelet-activating factor receptors. Behav Neural Biol, 1990, 54: 213～217

79 Clark G D, Happel L T, Zorumski C F, Bazan N G. Enhancement of hippocampal excitatory synaptic transmission by platelet-activating factor. Neuron, 1992, 9: 1211～1216

80 Matsushime H, Roussel M F, Ashmun R A, Sherr C J. Colony-stimulating factor 1 regulates novel cyclins

during the G1 phase of the cell cycle. Cell, 1991, 65 (4): 701－713

81 Won K A, Xiong Y, Beach D, Gilman M Z. Growth-regulated expression of D-type cyclin genes in human diploid fibroblasts. Proc Nati Acad Sci USA, 1992, 89 (20): 9910－9914

（万晶宏）

第十三章　肿瘤发生与发展的比较蛋白组研究

第一节　蛋白质组与肿瘤

蛋白质组是指细胞、组织或基因组所表达的全部蛋白质，是由 Wilkins 等于 1995 年首先提出[1,2]。蛋白质组研究具有观察由多基因事件引起的多蛋白质组分整体变化的独特优势。在人类所有疾病中，只有 2% 源于单基因缺陷（如基因替代或缺失），其余大部分是受遗传因素和环境因素综合作用的多基因事件[3]。对于癌症这一类复杂疾病，参与的基因更多。由于差异蛋白质组（功能蛋白质组）研究具有在整体水平上发掘和寻求潜在药物标靶以及适用于肿瘤分类、早期诊断、阻断（intervention）和治疗肿瘤标志物的能力，使得蛋白质组研究一提出便受到众多肿瘤研究者的青睐[4,5]。膀胱癌[6,7]、结肠癌[8]、乳腺癌[9]、肝癌[10]、食道癌[5,11]等均已进行不同深度的研究。传统研究肿瘤的单位都加大在肿瘤蛋白质组研究领域的投资力度。美国国家癌症研究所（NCI）资助密歇根医药大学 1000 万美元以进行在肺癌、结肠癌、肝癌、乳腺癌和卵巢癌蛋白质组领域的研究，其数据将和基因组及转录组数据结合，创建关于这些癌症的信息库。同时 NCI 还资助数百万美元用于 FDA 在组织蛋白质组领域的研究[12]。1999 年国家自然科学基金重大项目“蛋白质组以及蛋白质结构动态变化与其生物功能的研究”正式立项，其中包含“肿瘤发生与发展”、“肿瘤恶性标志”等领域的研究。2001 年蛋白质组研究正式被纳入国家重点基础研究规划项目《恶性肿瘤发生与发展的基础性研究》子项目。

第二节　肺癌转移的蛋白质组研究

肺癌作为威胁人类健康的重大疾病，每 8 个肺癌患者中有 7 个在初次诊断时就已经发生了转移，故大约 70% 的肺癌患者即便是清除了原发灶也会死于肿瘤的转移，而目前的化疗手段对肺癌的转移效果不大。因此，找到影响肿瘤转移的关键分子，然后针对性地设计药物抑制肿瘤的转移是目前治疗肺癌的最佳手段。但由于肿瘤的侵袭和转移的过程相当复杂，包括：① 肿瘤细胞间黏附能力降低，肿瘤细胞从原发部位脱落；② 肿瘤细胞在自身或间质细胞分泌的蛋白水解酶的作用下降解新生血管的基底膜和内皮，肿瘤细胞进而进入血液循环或淋巴系统；

③循环的肿瘤细胞在一些运动因子或归巢因子及受体的作用下，通过信号转导途径到达新的转移靶器官，并穿出血管，通过膜受体结合基质蛋白，刺激血管增生，形成新的转移灶。在肿瘤转移的每一步都受到了多个基因或蛋白的调控，有的已经过生物学功能的证实，如：上皮性钙黏着蛋白（E-cadherin）与胞内连环蛋白（catenin）的缺失，会导致胃、前列腺、直肠癌细胞的细胞间黏附作用，转移能力增加[13,14]；蛋白水解系统中起主要作用的有尿激酶纤溶酶原激活物（uPA）[15]、uPAR（uPA 受体）抑制剂 PAI、由肿瘤细胞和间质细胞分泌的基质金属蛋白酶（MMP）及抑制剂 TIMP 系统；在肿瘤细胞黏附、迁移过程中起重要作用的 integrin 等；有的还有待于进一步的研究。因此，要在众多的因素中找出起关键作用的蛋白分子，以此为依据设计相应的药物，蛋白质组技术的高通量、高灵敏度、高重复性的特点无疑会起到重要作用，而目前对肿瘤转移的蛋白质组研究，还有较大的空间。

一、肿瘤转移模型的建立

好的模型的建立是实验成功的关键，而目前的肿瘤转移模型主要通过以下方式建立：① 体内筛选实验可以从亲代的细胞中分离出具有不同转移能力的细胞株。Gemma 等[16]将亲代的肺腺癌细胞（PC9）反复接种到裸鼠体内后，把转移至肺的肿瘤组织反复培养后获得了具有高侵袭能力和不同黏附能力的高转移细胞株。Kozaki 等[17]通过两轮的裸鼠体内皮下接种后，将肺组织的转移结节进行体外的培养，进而将获得的细胞皮下再接种裸鼠，继续将自发性的淋巴结转移细胞培养后获得了具有高淋巴结转移能力的肺癌细胞株。② 体外筛选实验获得不同转移能力的肺癌细胞株。Chu 等[18]将肺腺癌细胞通过包被了人工基底膜的 Transwell Invasion Chamber 后，通过几轮筛选获得具有不同侵袭能力的细胞株。军科院的陆应麟教授将有限稀释后的单克隆细胞接种裸鼠皮下，获得了具有不同肺和淋巴结自发转移能力的肺巨细胞癌细胞株[19]。除了将具有不同转移和侵袭能力的细胞株作为肿瘤转移的蛋白质组研究对象外，还可采用肺癌患者的原发部位和转移灶的肿瘤组织进行比较蛋白质组分析，从而寻找出与肿瘤转移相关的蛋白，阐明转移机制。

二、肺癌转移模型的鉴定

为了保证肺癌转移模型的可靠性，需要对建立的转移模型进行确定。相关的体内和体外实验包括以下几方面：

(1) SCID 小鼠或裸鼠的体内成瘤及自发性转移实验

将肿瘤细胞通过皮下或尾静脉异种移植到小鼠体内，数周后解剖小鼠的内脏

器官（心、肝、肾、脾、肺、脑等）及淋巴结，石蜡包埋，切片，HE 染色，显微镜下观察肿瘤在小鼠体内的自发性转移情况。

（2）正位移植

将 1×10^7 的细胞通过插入在颈部皮下的支气管注射器针头直接接种细支气管泡腔，1 个月后解剖肺和纵隔，石蜡包埋切片，HE 染色，显微镜下观察纵隔淋巴结的转移情况。

（3）明胶酶谱分析

MMP-2（明胶酶 A、或 72kDaⅣ型胶原酶）、MMP-9（明胶酶 B 或 92kDaⅣ型胶原酶）属基质金属蛋白酶家族，是一类 Zn^{2+} 依赖的蛋白水解酶，二者在胞外基质和基底膜的降解过程中发挥重要作用，为肿瘤细胞的侵袭和转移提供可能。分析方法如下：接种对数生长期的细胞（1×10^5）于 24 孔板中，37℃，5% CO_2 培养 24h 后，用无血清培养基轻洗三次，加入 500μl 无血清培养基培养 24～48h，收获上清，离心，超滤浓缩后用作酶谱分析。于 10μl 上清中加入 10μl 非还原的样品缓冲液，上样，在含 0.1%明胶的 10%SDS-PAGE 上电泳。电泳结束后，蒸馏水中漂洗两次，2×10min；将胶浸入 2.5% Triton X-100 中洗脱 SDS，1h 内换液两次；将胶转入明胶酶解缓冲液（20mmol/L Tris-HCl，pH7.5；1μmol/L $ZnCl_2$；10mmol/L $CaCl_2$；150mmol/L NaCl）中，37℃反应 24h；0.5%考马斯亮蓝（30%甲醇、10%冰醋酸）染色，脱色至透明条带在蓝色背静下清晰为止。

（4）体外侵袭实验[20]

采用膜侵袭培养系统可反映细胞在体外的侵袭能力高低。将含 5×10^4～2×10^5 细胞的培养液加入 transwell chamber（Costar 公司）的上室，下室加入 NIH-3T3 细胞的条件培养基，上下室之间用包被了 Matrigel（50μg/室）的聚碳酸膜（6.5mm 直径，8μm 孔径）隔开，37℃，5% CO_2 培养 24～48h 后，PBS 洗膜内、外三次，用棉签轻擦去膜内层细胞，甲醇固定穿膜细胞 30min，HE 染色 10min，待干后将 Chamber 揭下并用树脂封片，显微镜下计数穿膜细胞，计算穿膜细胞占总细胞的百分率。如果为了检测细胞在体外的迁移能力，除了聚碳酸膜勿需包被及加入的细胞数改为 10^4 外，实验方法同体外的侵袭实验。对于细胞在体外的黏附实验来讲，实验步骤如下：将 7.5×10^3 细胞加入包被了 Matrigel/ collagen/ fibrolectin/laminin 的 96 孔板中，37℃，5% CO_2 培养 3.5h，洗去未黏附的细胞，显微镜下计数，计算黏附细胞占细胞总数的百分率。

（5）转移相关指标的检测

在肿瘤的转移过程中，受多个基因（癌基因和抑癌基因）的调控，如：上皮性钙黏着蛋白（E-cadherin）与胞内连环蛋白（catenin）、Ig 超家族、整合素、选择素在介导肿瘤细胞与胞外基质或内皮细胞间的黏附中发挥作用，促进肿瘤细胞穿越血管；纤溶酶原激活物（PA，包括 uPA 和 tPA）及纤溶酶原激活因子抑制

剂 PAI 系统、基质金属蛋白酶 MMP 及组织型基质金属蛋白酶抑制剂 TIMP 系统在胞外基质的降解过程中发挥重要作用；促进肿瘤细胞迁移运动的因子 HGF、EGF、IL-1、3、6、TGF 等；此外，VEGF、TGF-α、TNF-α、MMP、TIMP、PA、IL-8、bFGF 等促进微血管生成的因子与 IFN-α β、TGF-β、PAI、TIMP、endostatin、血管生成抑制素、血小板因子 4 等抑制微血管生成的因子之间的平衡在肿瘤的微血管生成机制方面也发挥着重要作用。通过 Western blot、RT-PCR 或 Northern blot 等各种手段对与肿瘤转移相关指标的检测，有助于阐明在肺癌转移过程中各种蛋白之间的相互关系，明确起主要作用的关键蛋白。

三、肿瘤转移的比较蛋白质组分析

近年来，蛋白质组技术在研究有关正常细胞的分化、增殖、异常转化、肿瘤的发生、发展方面进行了探索，涉及到了血液系统肿瘤、乳腺癌、结肠癌、膀胱癌、前列腺癌、肺癌、肾癌和神经母细胞瘤等，在寻找肿瘤早期诊断的生物标志物及新的分子治疗的药靶、评价药物的疗效方面提供了一些新的思路。但迄今为止，对有关肿瘤转移的研究都是比较零星分散的，应用蛋白质组技术探讨肿瘤转移发生的相关文献较少。Hirano 等[21]报道通过蛋白质组技术找到的 TA01/TA02 异构体不仅可作为原发性肺癌的早期诊断指标，还可作为与远端转移性肺癌鉴别的诊断标志。此外，Alaiya[4]通过比较蛋白质组技术分析了乳腺纤维瘤和不同组织类型、不同分化程度的临床乳腺癌标本，发现在浸润性肿瘤中出现了 PCNA、应激蛋白（HSP60、HSP90、calreticulin）的高表达及原肌球蛋白 1～3 的低表达，随后将原肌球蛋白 1 的 cDNA 导入细胞后，出现了转化表型的逆转。Harvey 等[22]找出了几个只在转移性的乳腺癌细胞株胞膜表达的蛋白，如：在细胞的恶性转化过程中发挥作用的真核细胞启动因子 3 的 P47 亚单位（eIF3）；在乳腺癌肿瘤转移中关系不太明确的 α-Synuclein，但有报道，γ- Synuclein 转染 MDA-MB-435 细胞后，细胞株的转移潜力提高；通过介导 MDA-MB-435 细胞同内皮细胞黏附而促进转移的 Galectin-1；通过抗氧化而起到保护细胞作用的巯基特异性的抗氧化蛋白（TSA）。但仍有多个在肿瘤转移中发挥作用的低丰度蛋白未能得到鉴定，需进一步分析。但至今为止，还未见有关肺癌转移的系统性的蛋白质组分析。故此在以下的章节将结合本实验室的工作讨论有关肺癌转移的蛋白质组研究。

1. 样品的制备

由于样品制备的好坏直接关系到 2-DE 的分离效果，因此一直以来，人们在样品准备方面不断进行着改进。但由于肿瘤组织本身的异质性（标本中癌细胞和非癌细胞的类型及比例不同；肿瘤内不同组织损伤、凋亡、坏死程度不同；肿瘤

不同部位细胞的基因改变导致蛋白表达不同），很难用一种溶解液溶解所有的蛋白，且经 2-DE 分离后的蛋白表达谱相关性较差。Alaiya 等[23]报道不同患者的卵巢肿瘤组织标本间的蛋白质组相关性仅为 0.54～0.72，同一肿瘤组织内不同损伤部位的相关性平均为 0.85。但如果仅仅采用肿瘤细胞株或组织匀浆标本作为研究对象，则很难反应出组织内特定细胞群的蛋白质组改变。如今，激光捕获微切割（laser capture microdissection，LCM）技术的诞生为组织标本的样品制备带来了新的曙光。由于 LCM 技术是在显微镜下采用激光对组织中某一特定类型的细胞进行切割，因此避免了组织中其他组分的干扰，所获细胞单纯，能较好地反映某一细胞群的蛋白质组成分改变。但该技术的一个最大缺点在于其通量低，往往需数天时间才能获得 2-DE 分析的足够样品，无法满足大规模蛋白质组研究的要求[24-26]。

此外，体液标本（如血、尿、脑脊液、滑液等）在肿瘤蛋白质组学的研究也偶有报道[27]，但因为体液中包含了来源于不同组织和病程的复杂蛋白成分，就使其蛋白质组研究受限。再者，血液和脑脊液中 99% 的成分是血清白蛋白和球蛋白，它们就掩盖了低丰度蛋白的检出。而低丰度蛋白往往具有重要的生理与病理意义，对于这些低丰度蛋白，近年来人们尝试通过亲合纯化及免疫沉淀结合 SDS-PAGE 的预分离技术富集具有共同特征的蛋白群，取得了较好效果，可分离更多的蛋白点[28]。

用作肿瘤转移的样品可以采用转移能力不同的细胞株，也可以采用肿瘤的原发灶与转移灶进行比较蛋白质组分析。本实验室以人肺巨细胞癌（PLA-801-D、C）高、低转移株作为研究对象，细胞在含 10% 的小牛血清的 RPMI 1640 中于 37℃、5%CO_2、饱和湿度条件下培养。待细胞培养至对数生长期，胰酶消化，清洗离心，收集 1×10^6 细胞悬于 60～100μl 的裂解液中（8-9 mol/L urea，4% CHAPS，40mmol/L Tris 和 65 mmol/L DTT，内含 10mmol/L PMSF）。之后，快速冻融 3～4 个循环，以 13 000r/min 离心 30min，样品储存于 －70℃，采用 Bradford 法对蛋白进行定量。

2. 制备型 2-DE 分析

（1）等电聚焦

通过 IPGphor 进行第一向等电聚焦（聚焦条件如表 13-1 所示）后，平衡，在 Protein II IPG conversion kit 上进行垂直 SDS-PAGE 第二向的分离。凝胶经过考马斯亮蓝染色之后，利用光密度仪扫描，PDQUEST 软件分析获取有差异的蛋白点。

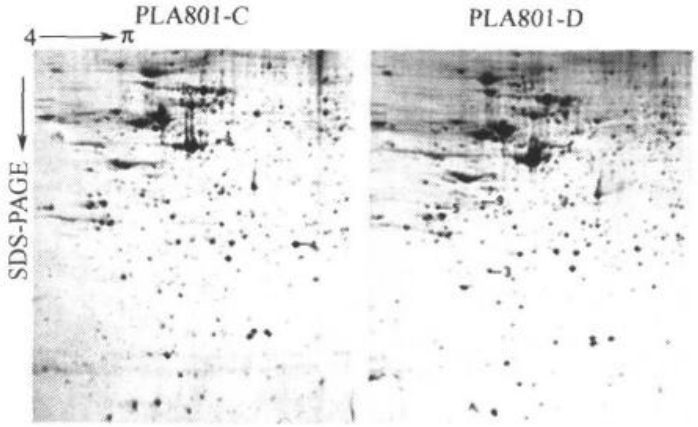


图 13-1　PLA801C、D 细胞株总蛋白的 2-DE 图谱
（pH4～7 线性，考马斯亮蓝染色）

表 13-1　第一相等电聚焦参数的选择

步骤	电压/V	电压-时间/Vh	时间/h	梯度类型
重泡涨	30	180	6	Step-n-hold
重泡涨	60	420	7	Step-n-hold
1	200	200	1	Step-n-hold
2	500	500	1	Step-n-hold
3	1000	750	1	Gradient
4	8000	64 000	8	Step-n-hold

(2) 平衡

将等电聚焦后的 IPG 胶条放入平衡液（1.5mol/L Tris 1.34ml，尿素 14.14g，甘油 12ml，SDS 0.8g，痕量溴酚蓝）40ml 中，于摇床上振荡 15min×2。其中，第一种平衡液中含 DTT 200mg，而第二种平衡液内由碘乙酰胺 0.9g 取代。

(3) 第二向垂直平板 SDS-PAGE

根据 IPG conversion kit 的体积，配制 12% 的均匀胶。将平衡后的 IPG 胶条移至凝胶的上方，用 0.5% 的琼脂糖电泳缓冲液封胶。电泳缓冲液为 Tris-Glycine-SDS，在 14℃时，以每一胶条 20mA 恒定电流电泳 30min 后，30mA 恒定电流电泳直至溴酚蓝前沿到达玻璃板下缘为止。

(4) 染色

电泳结束后，将待分析的凝胶放入固定液中固定 3h 或过夜；将凝胶放入考马斯亮蓝 R-250 染色液中染色，然后脱色至背景清晰。

(5) 图像分析

利用 GS-710 光密度仪，在 84.7μm 的分辨率下，对考马斯亮蓝 R-250 染色的凝胶进行扫描，利用 PDQUEST 6.1 软件进行图像分析。将经三次提取的三个

批次蛋白混合物，进行三次 2-DE 分析，获得了 9 张 2-DE 图谱。将 PLA801C 株细胞的 9 张 2-DE 图谱与 PLA801D 株细胞的 9 张 2-DE 图谱进行配比，根据图像分析结果，获取有 2 倍差异的蛋白点作为候选的鉴定蛋白质。

举例来讲，肺癌高、低转移株细胞的总蛋白采用 pH4～7 线性的 IPG 胶条，12%的 SDS-PAGE 电泳后，获得的 2-DE 图谱，经 PDQUEST 6.1 软件分析后，获得了有 2 倍量变的蛋白点，如 1＃号蛋白，在 PLA801 细胞株中高表达，可用作进一步的 MS 分析。

600μg 的总蛋白上样后，经 2-DE 分离，考马斯亮蓝染色鉴定，PDQUEST 软件分析后确定在 PLA80C 细胞中可检出 1024 个蛋白点，而在 PLA801D 细胞可检出 986 个蛋白点；且 PLA801C、PLA801D 细胞株间的 2-DE 相关性为 82%。

将三批样品分别进行 pH4～7 的 2-DE 分离后，PDQUEST 软件分析，将 PLA801C、D 细胞株的 2-DE 图谱进行配比，把有 2 倍量变的蛋白斑点作为待鉴定的差异候选蛋白。

将三批样品的 2-DE 图谱经 PDQUEST6.1 软件分析后，与参考图谱进行匹配，所获得的有 2 倍量变的蛋白点用黄色的图标在参考图谱上显示，红色圆圈内的蛋白点则为与之相匹配的样品 2-DE 图谱中的蛋白点，如图 13-2 所示，经软件分析可以得出红色圆圈内的蛋白点在 PLA801C、PLA801D 细胞间存在着 2 倍的量变，可作为差异候选蛋白。

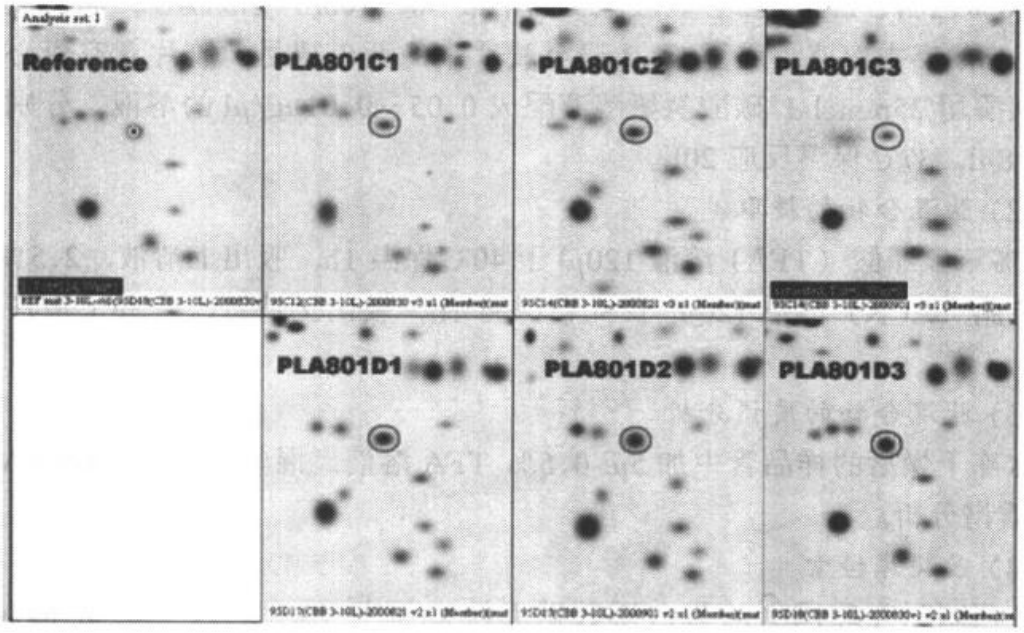


图 13-2　三批样品的 2-DE 图谱经 PDQUEST 软件分析配比后的结果
（局部放大图）

3. 肽质量指纹图谱鉴定候选的差异蛋白

经 2-DE 分离后的蛋白，可通过以下几种途径进行鉴定：①与已知数据库比较；②与已知蛋白平行电泳；③将蛋白转印至 PVDF 膜后进行 Edman 降解测序或内肽测序；④在真核细胞内瞬时表达；⑤质谱鉴定。在以上这几种鉴定方法中，以质谱技术应用最广。质谱作为一种快速、灵敏的蛋白鉴定技术，近几年来在蛋白质组技术领域得到了飞速发展。其中以 MALDI 和 ESI-MS/MS 结合四极杆、离子阱或飞行时间检测器等较为常用。前者适于高通量、大规模的蛋白质组分析；后者与液相色谱联用可鉴定复杂样品，准确度高。Gevaert 等[29]将经过 2-DE分离后的蛋白进行三步质谱鉴定，为我们提供了参考：①MALDI-TOF-PMF 分析；②若第一步失败，可将剩余的肽混合物进行 RP-HPLC 分离，再行 MALDI-PSD鉴定；③若前两步均失败，则需从头进行 ESI-MS/MS 分析。新近出现的另外一种质谱鉴定新技术 nanospray-LC-MS-MS 不仅可以增强检测的灵敏度（10^{-15}～10^{-18}mol），而且大大提高了鉴定效率（200～300 个蛋白/天），从而为蛋白质组的高通量分析提供了保障。对于临床标本而言，将 MS 技术与 immunowalking、免疫印迹、免疫组化等技术结合，更能反映复杂生物样品在体内的真实情况[5,6]。

(1) 蛋白斑点的脱色和胶上原位酶切

选取双向电泳凝胶上的候选蛋白斑点，用解剖刀片小心沿染色边缘切下蛋白点，并切成 1～2 mm^2 大小的胶片，另切一块无蛋白的 2D 凝胶作为空白对照。用含 50%乙腈、25mmol/L 碳酸氢氨的溶液 50～100μl 浸泡胶片，涡旋混合器振荡 20min，弃去溶液，重复 1～2 遍至胶片中的蓝色褪尽，胶片真空离心冷冻干燥。胰酶用 25mmol/L 碳酸氢氨溶液配成 0.05～0.01μg/μl 的溶液，分别加入样品 5～8μl，37℃保温反应 20h。

(2) 肽混合物的提取

5%三氟醋酸（TFA）溶液 120μl 于 40℃保温 1h，吸出上清液，2.5% TFA、50%乙腈（ACN）溶液 120μl 于 30℃保温 1h，吸出上清液，合并上清液冰冻干燥。

(3) 肽混合物的质谱分析

冰冻干燥后的样品管中加 5μl 0.5% TFA 溶液，混匀后取 1μl 进行 MALDI-TOF 质谱分析。

(4) 数据库检索

使用国际互联网上的蛋白质数据库检索程序（www.expasy.ch/tools/peptident.html）对肽质量指纹图 PMF 数据进行检索，从而鉴定该蛋白。以 PLA801D 细胞株中高表达的候选蛋白为例，该蛋白经酶解后 MALDI-TOF 分析所获得的 PMF 如图 13-3 所示：

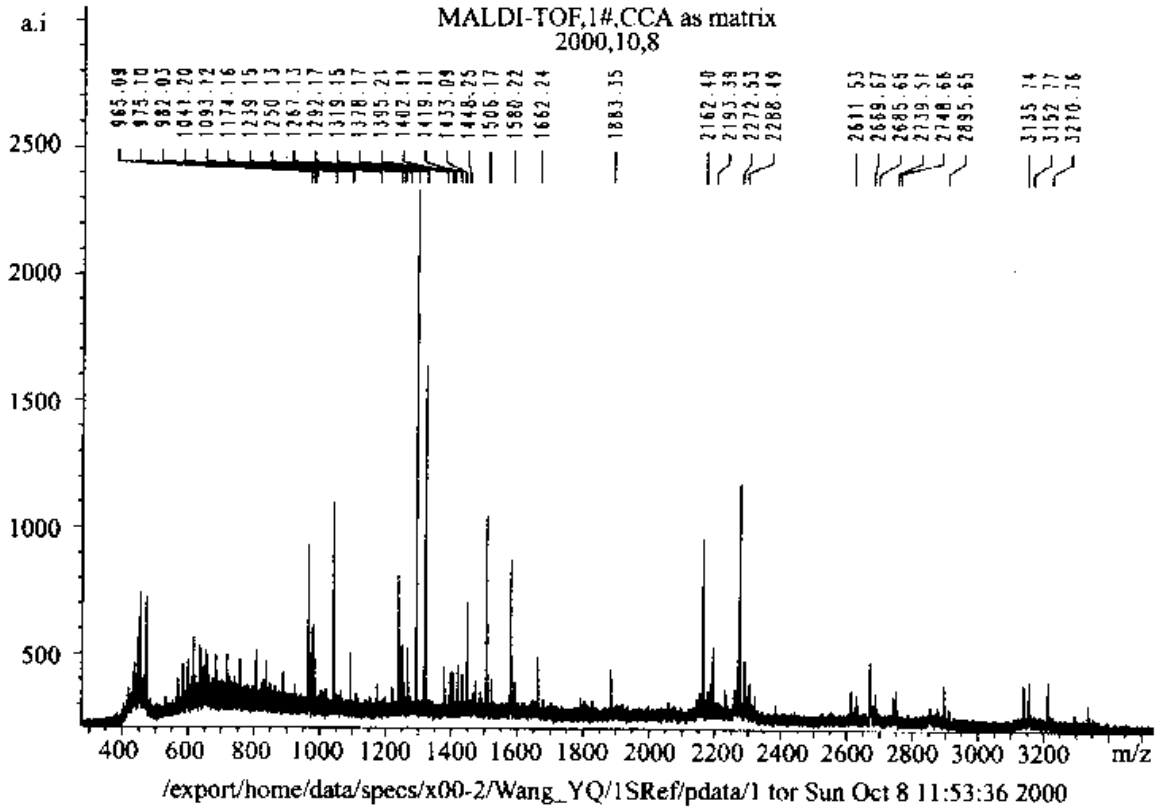


图 13-3　候选蛋白的肽质量指纹图谱

其肽质量数分别为：965.09、975.10、982.03、1041.20、1093.12、1174.16、1239.15、1250.13、1267.13、1292.17、1319.15、1378.17、1395.21、1402.11、1419.11、1433.09、1448.25、1506.17、1580.22、1662.24、1883.35、2669.67、2685.65、2739.51、2748.66、2895.65，进行数据库查询，查询条件如表 13-2 所列：

表 13-2　肽质量指纹谱数据库查寻参数设定

database	SWISS-PROT
Protein mass range	40 000-60 000
pI range	4.5～5.5
species to be searched	MAMMALIA
all peptide masses are	monoisotopic
Mass tolerance	1.0Da
enzyme	Trypsin
number of peptides required for match	4
Cystein is treated with	iodoacetamide

查询结果显示如表 13-3：

表 13-3 候选蛋白经 PeptIdent 数据库的查询结果

Score	# PeptIde matches	AC	ID	Description	pI	Mw
0.73	19	P05783	K1CR _ HUMAN	Keratin, type I cytoskeletal 18 (Cytokeratin 18) (K18) (CK 18). -Homo sapiens (Human).	5.34	47926.62
0.23	6	P05371	CLUS _ RAT _ 1	Chain 1: clusterin. -Rattus norvegicus (Rat).	5.46	49076.38
0.23	6	P05784	K1CR _ MOUSE	Keratin, type I cytoskeletal 18 (Cytokeratin 18) (Cytokeratin endo B) (Keratin D). - Mus musculus (Mouse).	5.22	47372.89
0.23	6	Q76013	K1M6 _ HUMAN	Keratin, type I cuticular HA6 (Hair keratin, type I HA6). - Homo sapiens (Human).	4.90	52247.07
0.23	6	P05383	KAPB _ PIG	cAMP-dependent protein kinase, beta-catalytic subunit (EC 2.7.1.37) (PKA C-beta) (Fragment). - Sus scrofa (Pig).	undef-ined	undef-ined
0.23	6	Q03924	Z117 _ HUMAN	Zinc finger protein 117 (Zinc finger protein HPF9) (Fragment). - Homo sapiens (Human).	undef-ined	undef-ined
0.19	5	P46660	AINX _ MOUSE	Alpha-internexin (Alpha-Inx) (66 kDa neurofilament protein) (Neurofilament-66) (NF-66). - Mus musculus (Mouse).	5.16	55869.98
0.19	5	P23565	AINX _ RAT	Alpha-internexin (Alpha-Inx). - Rattus norvegicus (Rat).	5.20	56115.38
0.19	5	O08715 (Isoform)	AKA1 _ MOUSE	SPLICE Isoform 5 of A kinase anchor protein 1, mitochondrial precursor (Protein kinase A anchoring protein 1) (PRKA1) (Dual specificity A-Kinase anchoring protein 1) (D-AKAP-1) (Spermatid A-kinase anchor protein) (S-AKAP). - Mus musculus (Mouse).	4.82	57969.54
0.19	5	P06727	APA4 _ HUMAN _ 1	CHAIN 1: APOLIPOPROTEIN A-IV. - Homo sapiens (Human).	5.18	43374.52

经 PeptIdent 检索结果表明，细胞骨架蛋白 18（Keratin, type I cytoskeletal 18, K18）(Cytokeratin 18, CK 18) 的肽质量指纹谱与 1 号蛋白斑点的质量肽谱一致，均为人源性的蛋白，得分为 0.73，序列覆盖率为 48%，且与位于第二位的蛋白得分（0.23）差异较大，故认为 1 号蛋白点应该是 CK18。

四、差异候选蛋白的进一步确证

对于在 PLA801C、D 细胞株中有差异的蛋白，MS 鉴定只能作为其中的一个有价值的指标，为了保证差异蛋白的可靠性，需采用其他的一些手段对差异蛋白进行进一步的验证，才能保证差异蛋白确实与肿瘤转移相关。Celis JE 实验室结

合蛋白质组技术和冰冻切片的免疫组化分析鉴定出了几个与膀胱鳞细胞癌分化程度相关的蛋白，为指导临床医生判断病情、采取治疗措施提供了依据[30]。O'Shaughnessy RF[31]等结合蛋白质组技术和 cDNA 阵列杂交技术证实：通过蛋白质组技术获得的差异表达蛋白与通过 cDNA 阵列杂交技术所获得的差异表达基因绝大部分相似，其中上皮性脂肪酸结合蛋白（PA-FABP）的高表达可作为人表皮移行扩增细胞的标志蛋白。

对于在肺癌高、低转移株细胞中有差异表达的候选蛋白，我们从两个方向对其进行了确证：① 在蛋白水平，用 Western blot 技术对候选的差异蛋白进行验证；② 在分子水平，采用半定量 RT-PCR 结合 Northern blot 技术对编码候选差异蛋白的 RNA 进行确证。以上面提到的蛋白为例，通过 2-DE 技术分离后，发现该蛋白在 PLA801C、D 株细胞间有 2 倍以上的量变，为了证实该蛋白在 C、D 株细胞间确有差异，我们进行了以下的实验：采用 RIPA提取液提取 C、D 株细胞的总蛋白，等量上样，SDS-PAGE 电泳后，将胶转移至 PVDF 膜上，封闭，免疫反应，ECL 显色，压片，显影、定影。结果如下：从 Western blot 的结果可以看出，CK18 在肺癌高、低转移株细胞间存在显著差异，CK18 在肺癌高转移株细胞中的表达增加。Chu 等观察到 CK18 表达阳性的肺癌细胞，其转移能力较 CK18 阴性的细胞强，与肿瘤细胞的迁移能力增强和在胞外的扩散能力降低有关[18]。对于一些没有相应抗体的差异蛋白，可采取从分子水平进行验证的方法。对于 RNA 水平的检测，可以采用定量 RT-PCR 及 Northern blot 的方法，但由于蛋白水平与 RNA 水平的不一致性以及蛋白存在翻译后的修饰等情况，因此对 RNA 水平的检测不能完全反应蛋白水平的情况。

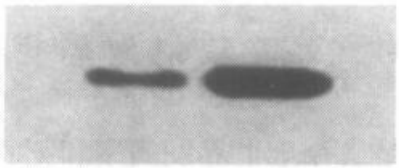


图 13-4 CK18 在 PLA801C、D 细胞株的差异表达（Western blot）

从 Western 结果（图 13-4）可以看出，CK18 在 PLA801C、D 细胞间存在显著差异，与 2-DE 的分析结果一致。

五、候选蛋白的功能探讨

通过蛋白质组技术可发现一些在肿瘤转移中未见报道的蛋白，为了观察这些蛋白在肿瘤转移中是否发挥了作用及通过何种途径起作用，无疑会给肿瘤的诊断和寻找治疗靶标提供线索。具体的策略可参照 C-FABP（cutaneous fatty acid-binding protein）的研究，Jing 等[32]通过系统性的差异显示技术找出了在前列腺和乳腺癌细胞中高表达的 C-FABP 基因，将 C-FABP 的 cDNA 导入良性的、非转移的大鼠乳腺上皮细胞株 Rama 37 后，获得的转染细胞接种同基因的 Wistar-

Furth 大鼠，结果发现有 6/26 的大鼠体内出现了肿瘤转移，而空载体转染的细胞接种后未出现肿瘤的转移，由此看来 C-FABP 基因与肿瘤的转移相关。随后的一系列实验证实 C-FABP 是通过促进 VEGF 的表达而增强肿瘤细胞的转移能力的[33]。随着研究的深入，将会发现一些新的蛋白可能与肿瘤的转移有关，而蛋白质组技术将会为肿瘤转移相关蛋白的发现、对肿瘤转移机制的研究起到重要作用。

第三节　辐射致癌相关蛋白质分子标志物的研究

肺癌是当今全世界发病率和死亡率增长最快的疾病，在欧美发达国家不论男女均占常见肿瘤的首位，在我国沿海城市男性中也占首位，在女性中居第二或第三位[34]。其主要病因来源于抽烟、大气粉尘及天然本底辐射等。原发性肺癌是环境因素与人体交互作用的结果，其发生、发展是多因素、多基因及多阶段的复杂过程。其中癌基因与抑癌基因的失衡是其发生、发展的主要机制。而仅仅在基因水平的研究是不够的，应该运用蛋白质组研究的技术体系，从整体水平上深入到基因功能的具体实施者——蛋白质来进行研究。目前关于肺癌蛋白质组水平上的研究非常有限，互联网及文献中也没有关于肺癌细胞系或组织的蛋白质表达谱发布。Okuzawa[35]等分析了 14 例不同组织病理类型的临床肺癌肿瘤样本，总结出一些蛋白的异常表达，并分析得到可能的肿瘤标志物 TA02，后经 Chuman[36]等鉴定为 aspartic protease napsin A。由于个体差异以及间质和正常组织的干扰，使得临床样本的可变性太大，很难分辨开有意义点和污染点，得到一个真正的肺癌蛋白质组表达谱。

癌症的一个诱发原因是暴露于电离辐射，辐射是一种已知的环境致癌因子，氡及其子体是人类本底辐射的最主要成分，其危害来自于沉积在呼吸道表面氡不断衰变产生的 α 粒子照射。目前氡致人类肺癌已成为研究热点，然而在临床或现场很难获得正常、癌前和癌组织，因此，建立体外支气管癌变模型有助于更方便地了解疾病发生、发展及其分子机制。本研究所采用的细胞为人支气管上皮细胞系 BEP2D，1991 年由 Harris 用克隆的全长人乳头瘤病毒（Human Papillomaviruses 18，HPV18）转染正常人支气管上皮细胞（NHBE）构建而成[37]，该细胞系在体外连续传代 100 次以上未发生衰老和死亡，接种裸鼠饲养 12 个月以上未观察到成瘤性，排除了融合诱导的鳞状分化，表明它们获得了永生化，是理想的肺癌研究模型。目前世界上只有两家研究所建立起完整的转化模型，有关该细胞系基因表达改变的研究已取得进展[38~40]。国家重点基础研究发展规划项目《恶性肿瘤发生与发展的基础性研究》子项目“肺癌前期病变发生发展及转轨的分子机制研究”小组用 α 粒子辐射诱导 BEP2D 细胞系，得到 RH 细胞系，传代培养至 20 代龄（RH20）时，裸鼠接种实验和形态学分析呈现出癌前病变形态，

表明该细胞系 20 代时处于癌前阶段；照后培养细胞至 35 代（RH35）时，病理学检查提示细胞已经开始急性转化，为高分化鳞癌，表明 35 代细胞已实现癌前到癌细胞的转化（如图 13-5 所示）。原代、20 代、35 代 RH 细胞系的获得为研究肺癌早期病变和转轨的分子机制提供了一个良好的模型。利用双向电泳、图像分析和质谱鉴定等手段，我们对 RH0、RH20、RH35 三代细胞表达图谱进行差异蛋白质组分析，获得了部分差异蛋白质点的信息。

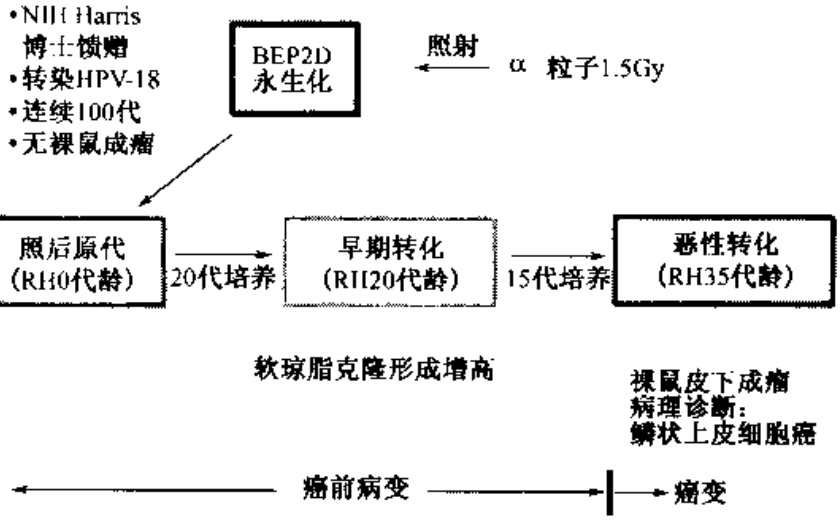


图 13-5　BEP2D 细胞恶性转化模型的建立

一、细胞培养与样品提取

永生化的人支气管上皮细胞系 BEP2D 来自于 Harris CC 教授的惠赠[37]。将该细胞在 37℃下，5% CO_2 浓度的培养箱中，用 HLC-8 无血清培养基培养。当细胞达到指数增长期后，胰酶消化细胞并用 PBS 缓冲液洗涤三遍。2×10^6 个细胞中加入 100μl 的裂解缓冲液（8 mol/L 尿素，4% CHAPS，40mmol/L Tris），在液氮中冻融几次后，加入 20μl RNA 酶（20μg/μl）。放置 15min，4℃下 13 000 r/min 离心 30min，吸取上清，贮存－70℃备用。蛋白定量采用 Bradford 法[41]。

二、双向电泳分离蛋白质

使用一向安法玛西亚公司 IPGphor 等电聚焦仪，二向 BIO-RAD 公司 PROTEAN II Cell，实验流程、程序设定参见本章第二节，分析胶的银染和制备胶考马斯亮蓝染色方法参见第四章。

三、图像分析

银染后分析胶利用 ImageScanner 光学扫描仪，在 256 色灰度和 300dpi 的分辨率下，对银染的凝胶进行扫描。在 Pentium Ⅲ 500 微处理器（Dell，128MB 内存）下运行，分析蛋白图谱。凝胶点的检测（Spot detection）、背景消减（Background subtraction）、相对分子质量和等电点校正、归一化处理（Volume normalization）和凝胶配比（Gel match）分析用 ImageMaster 2D Elite 3.01（Amersham Pharmacia）来完成。选择其中的一张凝胶图像作为参考胶（Reference gel），其他的图谱与之进行配比，并选择一些匹配点（User Seeds），以增强两块胶的自动配比能力。蛋白质斑点的量被定义为构成这个点所有象素点强度值的总和，为了较准确的反映蛋白质斑点量的变化，将整块胶中每个点的量（Vol）累加，每一个蛋白点的量除以累加值称为该蛋白质点占胶内所有蛋白质总量的百分含量（%Vol）。数据分析以该%Vol 为基准，在 Microsoft Excel 97 中进行，调用统计检验为 Student's t-test。

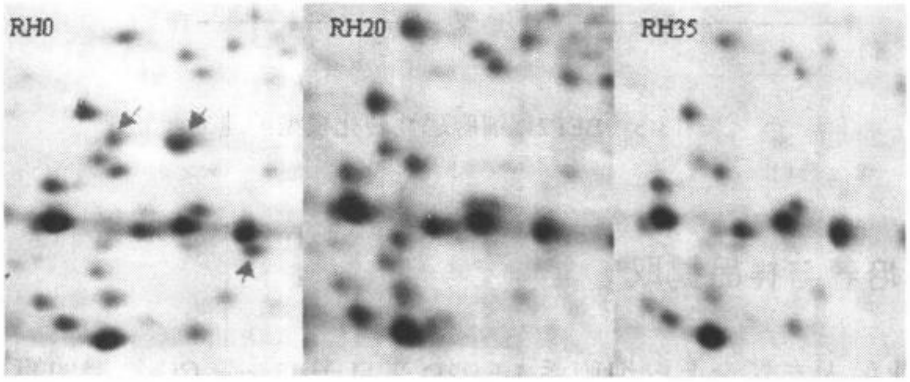


图 13-6　只在 RH0 代细胞系表达的三个蛋白质点

利用 ImageMaster 双向电泳图谱分析软件，分析结果显示有 3 个点只在原代细胞中表达（图 13-6），而 20 代和 35 代细胞表达并无显著差异。蛋白质表达量的差异运用 Student's t-test 进行显著性分析（$P<0.05$）。结果显示原代和 20 代之间有 43 个蛋白质点表达量发生变化，其中 22 个蛋白质表达量下降，21 个点的表达量上升（图 13-7）。20 代至 35 代的分析表明，有 66 个蛋白质点的表达量发生了变化，其中 53 个蛋白质表达量下降，只有 13 个蛋白质点表达量上升（图 13-8）。在 20 代和 35 代均发生表达变化的点有 18 个。

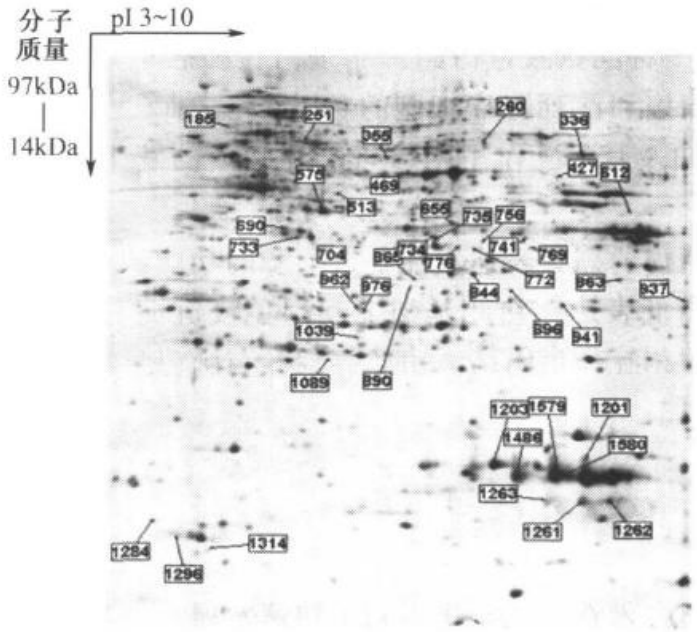


图 13-7　RH0 与 RH20 双向电泳差异表达蛋白质点

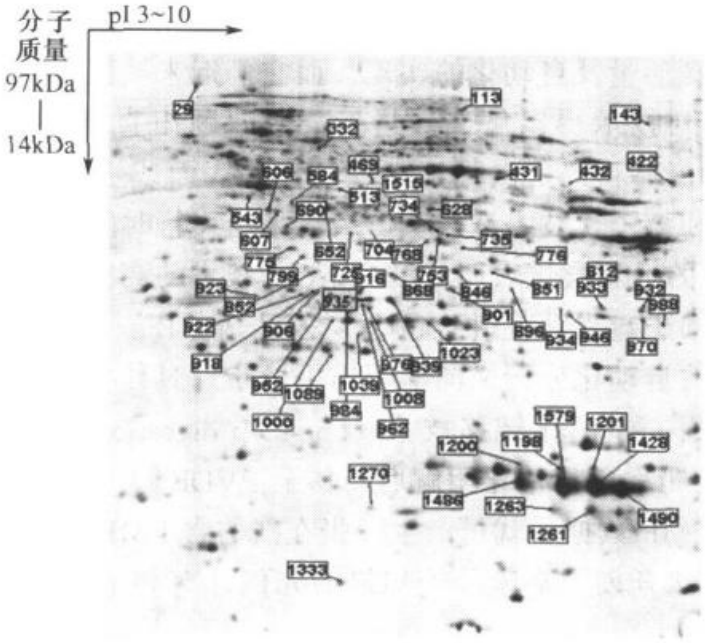


图 13-8　RH20 与 RH35 双向电泳差异表达蛋白质点

四、差异蛋白质点的指认

运用肽质量指纹谱技术和基质辅助激光解吸附飞行时间质谱仪源后衰变（post source decay，PSD）功能，对部分差异表达的胶上蛋白质进行酶切鉴定。如点 935 和点 939 经鉴定为高移动性家族蛋白 1（high mobility group protein 1，

HMG-1)，它们只在 RH 原代细胞表达而在 20 代和 35 代不表达。此外，我们发现 Maspin 蛋白，一种丝氨酸蛋白酶抑制剂，在肺癌早期转化过程中低表达，该蛋白可以抑制乳腺癌和前列腺癌细胞的移动和侵袭[42,43]。从对不同对象的研究来看，Maspin 蛋白及其相互作用网络可能在肿瘤的发生发展过程中扮演着重要角色[44~46]。

质谱鉴定所获得的差异表达点需要进一步的功能认证，如应用免疫印迹技术来确认蛋白质的高低表达，在此基础上利用免疫沉淀等技术，以钓取相关蛋白质，获得蛋白质的相互作用网络，由此寻找参与肺癌早期病变与发生发展的蛋白质（群）。

五、展望

尽管蛋白质组技术在近几年里得到了快速发展，但在样品制备、鉴定等方面还存在不足，如：对低丰度蛋白、极碱性蛋白、高相对分子质量蛋白的分离及对蛋白翻译后修饰的鉴定等，以及该技术在重复性、高通量、自动化等方面的不足，使得蛋白质组研究的技术平台目前还处于发展完善阶段。目前蛋白质组技术的主要瓶颈在于高通量及自动化的实现，而它们近来被证明有望通过逐渐发展完善的蛋白质微阵列技术得以解决。最近，两课题组直接将肿瘤组织标本或经 LCM 技术获取的样品粗提蛋白加于固相色谱柱表面或蛋白芯片阵列上，通过表面增强激光解离和离子化技术洗脱蛋白后，采用飞行时间质谱直接鉴定差异表达蛋白[47,48]。他们的最新尝试表明，结合 LCM 技术的高密度蛋白阵列、抗体阵列及小分子阵列有助于建立恶性肿瘤的蛋白表达谱。Binz 等[49]提出的分子扫描技术使得蛋白质组的自动化分析又向前迈出了一步。具体方法是将样品经 2-DE 微型胶分离后，采用一步消化转移技术（one-step digestion transfer，OSDT）技术将 2-DE 胶上所有蛋白平行消化酶解后转移至 PVDF 膜上，直接经 MALDI-TOF-MS 扫描，获取 PMF 数据后就可产生一张在线的含 PMF 数据和蛋白鉴定结果的 2-DE 图谱。该方法开创了临床分子扫描的先例，有利于肿瘤的临床诊断与预后评价。

参考文献

1 Wasinger V C, Cordwell S J, Cerpa-Poljak A, Yan J X, Gooley A A, Wilkins M R, Duncan M W, Harris R, Williams K L, Humphery-Smith I. Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis, 1995, 16 (7): 1090～1094

2 Wilkins M R, Sanchez J C, Gooley A A, Appel R D, Humphery-Smith I, Hochstrasser D F, Williams K L. Progress with proteome projects: Why all proteins expressed by a genome should be identified and how to do it. Biotech Gen Eng Rev, 1996, 13: 19～50

3 Strohman R. Epigenesis: the missing beat in biotechnology? Biotechnology (N Y), 1994, 12 (2): 156～164

4 Alaiya A A, Franzén B, Auer G, Linder S. Cancer proteomics: From identification of novel markers to creation of artifical learning models for rumor classification. Electrophoresis, 2000, 21 (6): 1210～1217

5 Emmert-Buck M R, Gillespie J W, Paweletz C P, Ornstein D K, Basrur V, Appella E, Wang Q H, Huang J, Hu N, Taylor P, Petricoin E F 3rd. An approach to proteomic analysis of human tumors. Mol Carcinog, 2000, 27 (3): 158～165

6 Celis J E, Wolf H, Ostergaard M. Bladder squamous cell carcinoma biomarkers derived from proteomics. Electrophoresis, 2000, 21 (11): 2115～2121

7 Ostergaard M, Rasmussen H H, Nielsen H V, Vorum H, Orntoft T F, Wolf H, Celis J E. Proteome profiling of bladder squamous cell carcinomas: identification of markers that define their degree of differentiation. Cancer Res, 1997, 57 (18): 4111～4117

8 Simpson R J, Connolly L M, Eddes J S, Pereira J J, Moritz R L, Reid G E. Proteomic analysis of the human colon carcinoma cell line (LIM 1215): development of a membrane protein database. Electrophoresis, 2000, 21 (9): 1707～1732

9 Vercoutter-Edouart A S, Czeszak X, Crepin M, Lemoine J, Boilly B, Le Bourhis X, Peyrat J P, Hondermarck H. Proteomic detection of changes in protein synthesis induced by fibroblast growth factor-2 in MCF-7 human breast cancer cells. Exp Cell Res, 2001, 262 (1): 59～68

10 曾嵘，俞利荣，邵晓霞，夏其昌. 电喷雾离子阱质谱法鉴定人肝癌细胞双向凝胶电泳胶内蛋白质. 科学通报，2000，45 (6)：592～598

11 夏书华，胡海，胡莉萍，徐昕，蔡岩，韩亚玲，陈宝生，魏芳，应万涛，钱小红，蔡有余，沈岩，吴旻，王明荣. 食管鳞状上皮癌中表达失调蛋白的研究. 癌症，2002，21 (1)：11～15

12 Abbott A. A post-genomic challenge: Learning to read patterns of protein synthesis. Nature, 1999, 402: 715～720

13 Cheng L, Nagabhushan M, Pretlow T P, Amini S B. Expression of E-cadherin in primaryand metastastic prastate cancer. Am J pathol, 1996, 148 (5): 1375～1378

14 Rimm D L, Sinard J H, Morrow J S. Reduced alpha-catenin and E-cadherin expression in breast cancer. Lab Invest, 1995, 72 (5): 491～493

15 Frandsen T L, Holst-Hansen C, Nielsen B S, Christensen I J, Nyengaard J R, Carmeliet P, Brunner N. Direct evidence of the importance of stromal urokinase plasminogen activator (uPA) in the growth of an experimental human breast cancer using a combined uPA gene-disrupted and immunodeficient xenograft model. Cancer Res, 2001, 61 (2): 532～537

16 Gemma A, Takenaka K, Hosoya Y, Matuda K, Seike M, Kurimoto F, Ono Y, Uematsu K, Takeda Y, Hibino S, Yoshimura A, Shibuya M, Kudoh S. Altered expression of several genes in highly metastatic subpopulations of a human pulmonary adenocarcinoma cell line. Eur J Cancer, 2001, 37 (12): 1554～1561

17 Kozaki K, Miyaishi O, Tsukamoto T, Tatematsu Y, Hida T, Takahashi T, Takahashi T. Establishment and characterization of a human lung cancer cell line NCI-H460-LNM35 with consistent lymphogenous metastasis via both subcutaneous and orthotopic propagation. Cancer Research, 2000, 60 (9): 2535～2540

18 Chu Y W, Yang P C, Yang S C, Shyu Y C, Hendrix M J C, Wu R, Wu C W. Selection of invasive and metastatic subpopulations from a human lung adenocarcinoma cell line. Am J Respir Cell Mol Biol, 1997, 17 (3): 353～360

19 陆应麟，黄靖香，李向红，李红芬，陈乐真，李维华. 人肺巨细胞癌 PLA-801 母系细胞系及其克隆化细胞株 (A、C、D、E) 裸鼠皮下种植后自发转移特性的研究. 中华肿瘤杂志，1989，11 (1)：3～7

20 Bjornland K, Winberg J O, Odegaard O T, Hovig E, Loennechen T, Aasen A O, Fodstad O, Maelandsmo G M. S100A4 involvement in metastasis: deregulation of matrix metalloproteinases and tissue inhibitors of

matrix metalloproteinases in osteosarcoma cells transfected with an *anti*-S100A4 *ribozyme*. Cancer research, 1999, 59 (18): 4702－4708

21 Hirano T, Fujioka K, Franzen B, Okuzawa K, Uryn K, Shibanuma H, Numata K, Konaka C, Ibihara Y, Takahashi M, Kato H, Auer G. Relationship between TA01 and TA02 polypeptides associated with lung adenocarcinoma and histocytological features. Br J Cancer, 1997, 75 (7): 978－985

22 Harvey S, Zhang Y, Landry F, Miller C, Smith J W. Insights into a plasma membrane signature. Physiol Genomics, 2001, 5 (3): 129－136

23 Alaiya A A, Franzen B, Moberger B, Silfversward C, Linder S, Auer G. Two-dimensional gel analysis of protein expression in ovarian tumors shows a low degree of intratumoral heterogeneity. Electrophoresis, 1999, 20 (4-5): 1039－1046

24 Ornstein D K, Gillespie J W, Paweletz C P, Duray P H, Herring J, Vocke C D, Topalian S L, Bostwick D G, Linehan W M, Petricoin E F 3rd, Emmert-Buck M R. Proteomic analysis of laser capture microdissected human prostate cancer and in vitro prostate cell lines. Electrophoresis, 2000, 21 (11): 2235～2242

25 Simone N L, Paweletz C P, Charboneau L, Petricoin E F 3rd, Liotta L A. Laser capture microssection: beyond functional genomics to proteomics. Mol Diagn, 2000, 5 (4): 301～307

26 Gillespie J W, Ahram M, Best C J, Swalwell J J, Krizman D B, Petricoin E F, Liotta L A, Emmert-Buck M R. The role of tissue microdissection in cancer research. Cancer J, 2001, 7 (1): 32－39

27 Kennedy S. Proteomic profiling from human samples: the body fluid alternative. Toxicol. Lett, 2001, 120 (1-3): 379－384

28 Corthals G L, Wasinger V C, Hochstrasser D F, Sanchez J C. The dynamic range of protein expression: A challenge for proteomic research. Electrophoresis, 2000, 21 (6): 1104－1115

29 Gevaert K, Vandekerckhove J. Protein identification methods in proteomics. Electrophoresis, 2000, 21 (6): 1145－1154

30 Celis J E, Celis P, Ostergaard M, Basse B, Lauridsen J B, Ratz G, Rasmussen H H, Orntoft T F, Hein B, Wolf H, Celis A. Proteomics and immunohistochemistry define some of the steps involved in the squamous differentiation of the bladder transitional epithelium: a novel strategy for identifying metaplastic lesions. Cancer Res, 1999, 59 (12): 3003－3009

31 O'Shaughnessy R F, Seery J P, Celis J E, Frischauf A, Watt F M. PA-FABP, a novel marker of human epidermal transit amplifying cells revealed by 2D protein gel electrophoresis and cDNA array hybridisation. FEBS Lett, 2000, 486 (2): 149－154

32 Jing C, Beesley C, Foster C S, Chen H, Rudland P S, West D C, Fujii H, Smith P H, Ke Y. Human cutaneous fatty acid-binding protein induces metastasis by up-regulating the expression of vascular endothelial growth factor gene in rat Rama 37 model cells. Cancer Res, 2001, 61 (11): 4357－4364

33 Jing C, Beesley C, Foster C S, Rudland P S, Fujii H, Ono T, Chen H, Smith P H, Ke Y. Identification of the messenger RNA for human cutaneous fatty acid-binding protein as a metastasis inducer. Cancer Res, 2000, 60 (9): 2390－2398

34 李连弟，鲁凤珠，张思维，牧人，孙秀娣，皇甫小梅，孙杰，周有尚，欧阳宁慧，饶克勤，陈育德，孙爱明，薛志福，夏毅. 中国恶性肿瘤死亡率 20 年变化趋势和近期预测分析. 中华肿瘤杂志，1997, 19 (1): 3－9

35 Okuzawa K, Franzen B, Lindholm J, Linder S, Hirano T, Bergman T, Ebihara Y, Kato H, Auer G. Characterization of gene expression in clinical lung cancer materials by two-dimensional polyacrylamide electrophoresis. Electrophoresis, 1994, 15 (3－4): 382－390

36 Chuman Y, Bergman A, Ueno T, Saito S, Sakaguchi K, Alaiya A A, Franzen B, Bergman T, Arnott D,

Auer G, Appella E, Jornvall H, Linder S, Napsin A. A member of the aspartic protease family, is abundantly expressed in normal lung and kidney tissue and is expressed in lung adenocarcinomas. FEBS Lett, 1999, 462 (1－2): 129－134

37 Willey J C, Broussoud A, Sleemi A, Bennett W P, Cerutti P, Harris C C. Immortalization of normal human bronchial epithelial cells by human papillomaviruses 16 or 18. Cancer Res, 1991, 51 (19): 5370～5377

38 楼铁柱，项小琼，吴德昌. α粒子诱发 BEP2D 细胞转化过程中染色体数目的变化. 癌变·畸变·突变, 2000, 12 (3): 172－173

39 Weaver D A, Hei T K, Hukku B, McRaven J A, Willey J C. Cytogenetic and molecular genetic analysis of tumorigenic human bronchial epithelial cells induced by radon alpha particles. Carcinogenesis, 1997, 18 (6): 1251－1257

40 Piao C Q, Willey J C, Hei T K. Alterations of p53 in tumorigenic human bronchial epithelial cells correlate with metastatic potential. Carcinogenesis, 1999, 20 (8): 1529－1533

41 Bradford M M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem, 1976, 72: 248－254

42 Sager R, Sheng S, Pemberton P, Hendrix M J. Maspin: A tumor suppressing serpin. Adv Exp Med Biol, 1997, 425: 77－88

43 Zou Z, Anisowicz A, Hendrix M J, Thor A, Neveu M, Sheng S, Rafidi K, Seftor E, Sager R. Maspin, a serpin with tumou-suppressing activity in human mammary epithelial cells. Science, 1994, 263 (5146): 526－529

44 Maass N, Hojo T, Zhang M, Sager R, Jonat W, Nagasaki K. Maspin-a novel protease inhibitor with tumor-suppressing activity in breast cancer. Acta Oncol, 2000, 39 (8): 931－934

45 Zhang M, Shi Y, Magit D, Furth P A, Sager R. Reduced mammary tumor progression in WAP-TAg/WAP-maspin bitransgenic mice. Oncogene, 2000, 19 (52): 6053－6058

46 Xia W, Lau Y K, Hu M C, Li L, Johnston D A, Sheng S J, El-Naggar A, Hung M C. High tumoral maspin expression is associated with improved survival of patients with oral squamous cell carcinoma. Oncogene, 2000, 19 (20): 2398－2403

47 Figeys D, Pinto D. Proteomics on a chip: Promising developments. Electrophoresis, 2001, 22 (2): 208－216

48 von Eggeling F, Davies H, Lomas L, Fiedler W, Junker K, Claussen U, Ernst G. Tissue specific microsection coupled with protein chip array technologies: application in cancer research. Biotechniques, 2000, 29 (5): 1066－1070

49 Binz P A, Muller M, Walther D, Bienvenut W V, Gras R, Hoogland C, Bouchet G, Gasteiger E, Fabbretti R, Gay S, Palagi P, Wilkins M R, Rouge V, Tonella L, Paesano S, Rossellat G, Karmime A, Bairoch A, Sanchez J C, Appel R D, Hochstrasser D F. A molecular scanner to automate proteomic research and to display proteome images. Anal Chem, 1999, 71 (21): 4981－4988

（蒋代凤　应万涛）

第十四章　细胞凋亡的蛋白质组研究

第一节　细胞凋亡的信号传导

细胞凋亡是多细胞有机体调节发育的一个重要特征。最初，从形态学特征定义凋亡，表现为细胞皱缩、质膜成泡状和染色质凝聚。凋亡细胞无炎性反应，最终形成凋亡小体[1]。与细胞分裂和迁移一样重要，凋亡（apoptosis）是由一组或者促进细胞凋亡或者促进细胞生存的基因紧密调控，决定细胞的数目和有机体的组织大小，因此在保持细胞内的稳态中发挥了重要的作用。不适的凋亡调节将导致各种疾病，包括癌症、自身免疫失调、病毒感染、AIDS、神经退行性病变和心肌梗塞等[2]。在过去的十多年中，其分子机制的研究揭示：凋亡大致要经历触发、调控和执行以及最后死亡特征的出现三个阶段。其中，细胞核内各种转录因子、细胞器线粒体内 Bcl-2 家族蛋白、细胞膜上各种功能复合体的募集（如 TRADD、FADD、CARD 等）、细胞质内系列半胱氨酸蛋白酶的级联放大、第二信使、细胞周期蛋白、骨架蛋白、结构蛋白以及内质网和（或）线粒体成分均参与凋亡形成，而其他与细胞生长、分化、增殖等密切相关的信号转导通路，如涉及对细胞的生存发挥重要作用的蛋白磷酸化的酶，c-Jun 激酶、p38 激酶和 cdc2 激酶皆与凋亡信号转导有联系[3]。现在有证据表明：其中两种凋亡的信号转导通路，即死亡受体信号通路和死亡受体非依赖（或者称为线粒体）传导通路，在细胞的一定状态下，会发生“对话”[4]。并且，依据有机体所处的环境，细胞个体还对于死亡和生存两种信号进行整合，来决定其命运。

一、细胞凋亡相关的几类信号分子

1. 细胞膜上分子支架的募集

位于细胞膜上的死亡复合物各种成分间的同型相互作用（homotypic interaction）对于凋亡信号是很关键的。它依附于分子支架（molecular scaffold）[5]，在细胞膜上通过将信号分子组装成功能复合体而提供局域性结构。它沿着信号转导通路，提供了类似“流水线”的方式。在凋亡分子中存在着三种类型的蛋白质相互作用结构域（interaction domain）：死亡结构域（death domain，DD）、死亡效应结构域（death effector domain，DED）和 caspase 募集结构域（caspase-recruitment domain，CARD）。DD 位于凋亡途径的上游，如死亡受体（CD95、TNFR

和 DR3）及募集这些受体的分子（FADD、TRADD 和 RIP）。而 DED、CARD 则负责募集与接头蛋白（adaptor protein）形成特定死亡复合物的长 N 端前结构域（prodomain），三者结构很相似[6]。其他相互作用的结构域，例如：与 RING、TRAF 结构域相关的富含半胱氨酸的基序（cysteine-rich motif associated with RING and TRAF domains, CART），伤亡白介素受体结构域（toll-interleukin-receptor domain, TIR）以及 MATH（meprin and the TRAF-homology domain）相互之间，与 DD-DED-CARD 间在结构上无关。而接头蛋白的设计原理是连接两个不同的相互作用结构域，亦通过同源相互作用来实现[7,8]。

DD 最初是由 Tarataglia 等人在研究肿瘤坏死因子受体（tumor necrosis factor receptor, TNFR）的缺失突变中首次提出[9]。位于蛋白质胞内区 C 端的 80 个氨基酸，负责产生细胞死亡信号。TNFR 超家族在哺乳细胞中迄今为止已有 20 多个成员被鉴定。这些蛋白在细胞外配体结合区域和细胞内效应区域均存在同源性。这些受体通过蛋白与蛋白间的相互作用来传递死亡信号。死亡结构域如 TRADD、FADD/MORT-1 以及接头蛋白 DAXX 给我们提供了与受体超家族相关的信号转导机制的新理解[10]。TRADD 是一个与其他已知数据库中蛋白质几乎没有同源性的基因产物，是与 TNFR1 相关的蛋白质。缺失突变研究表明，这种相互作用是通过 TRADD 的 C 端 111 个氨基酸 DD 来实现。TRADD 的 DD 亦调节自身相关的凋亡，其过表达可诱导凋亡，如细胞形态的改变和 DNA 断裂[11]。FADD/MORT-1 是另一个含有 DD 的蛋白，可广泛表达，且与 TNFR1 相互作用[12]。数据库搜索的结果表明 FADD 的 DD 与 TNFR1 和 Fas 有 25%～30%的同源性。与 TRADD 相似，FADD/MORT-1 的过表达可以诱导凋亡，但其效应区域在蛋白的 N 端。而接头蛋白 DAXX 通过其 C 端与 Fas 的 DD 结合[13]。DAXX 的过表达可以增强 Fas 介导的凋亡和激活 JNK 途径。此外，RIP（receptor-interacting protein）[14]是分子质量为 74kDa 的蛋白质，可与 Fas 和 TNFR1 结合，其 C 端是 DD，N 端与蛋白激酶具有同源性。RIP 在细胞内的过表达也可引起凋亡。CARDIAK/RIP2 则属于另一个新的 RIP 样激酶[15]。近来通过酵母双杂交和体外 GST pull-down 的鉴定表明，属于 $Ca^{(2+)}$-结合蛋白家族成员的凋亡相关基因 2（ALG-2），可以与 Fas 发生相互作用，表明 ALG-2 通过与 Fas 相互作用与凋亡具有直接的分子联系[16]，但其作用机制仍需进一步探讨。

2．胞质内的 caspase 的执行

在细胞质的均一混合弥散体系，存在着一系列蛋白酶 caspase，与凋亡“执行”密切相关，通过 caspase 的级联放大发挥作用。与线虫 Ced-3 和哺乳细胞白细胞介素 1-β 转换酶（IL-1β-converting enzyme, ICE）相似，caspase 为一高度保守的半胱氨酸家族蛋白酶，对含天冬氨酸残基的底物有特异性的酶切位点[17]。它在诸多细胞中存在，以单链酶原的形式位于胞质一侧。可通过两种方式激活：

自行断裂激活，或为另一个 caspase 所激活。两种情况皆产生有活性的异二聚体，由 10kDa 和 20kDa 的亚单位组成。其晶体结构显示，亚单位形成带有两个活性位点的异源四聚体。目前，凋亡中的 caspase 分为两类：启动者和执行者。caspase 存在长或短的 N 端前结构域。长前结构域（调节子）可断裂和激活短前结构域（效应子）。执行者的精确顺序和其他 caspase 的位置仍值得争论，但至少在 Fas 通路，死亡信号是通过顺序激活的。

caspase 与 caspase 相互作用在凋亡中是非常重要的，在凋亡过程中最初的 caspase 激活一定牵涉某些未知因子。目前，已知 caspase 酶原的聚集是导致 caspase 激活的重要一步，首先，由某些 caspase 前结构域的蛋白相互作用结构域与接头分子结合导致 caspase 酶原聚集，随后，通过下游的 caspase 的断裂和激活来进行级联放大凋亡信号。例如：caspase8、10 酶原包括两个串联死亡效应结构域（DED），而 caspase1、2、4、5 和 9 酶原含有 CARD，两者皆与接头蛋白相同的结构域结合。CARD、DED 具有与 DD 相似的结构，即由 6 个紧密包裹的、两性的、反向平行的 α 螺旋组成[18]。由此，导致凋亡顺序激活可概括为：① 上游信号导致的构象变化允许接头分子募集到死亡复合体；② 含特定结构域的 caspase 募集；③ 一旦出现邻近效应，caspase 酶原分子形成寡聚化过渡态，允许进行分子间或分子内的催化和激活。

由“执行者”caspase 断裂的蛋白，涉及 RNA 剪接、DNA 修复、细胞和核的骨架蛋白[19]。在体外凋亡实验中，由 caspase 断裂的第一个蛋白是 PARP（poly ADP-ribose polymerase），这是在某种极端状态下的底物，在体内可能是 caspase3、7。其中，caspase3 不但可断裂 DFF45/ICAD，形成特征性的染色质浓缩和断裂[20]，断裂核纤层，破坏核膜；而且还可断裂血影蛋白、肌动蛋白和 PAK2，使细胞膜和骨架的形态学改变。而“启动者”caspase 的作用位点却不同。其中，caspase9 可被 Bcl-2 家族成员在线粒体激活，负责断裂和激活 caspase3，而 caspase8 是 CD95/Fas 信号转导通路上的第一个酶。此外，caspase10 可启动 Fas 和 TNF 相关的配体-受体介导的凋亡，并不依赖于 caspase8。虽然 caspase8 和 10 的死亡效应结构域都可以与 FADD 的死亡效应结构域相互作用，但是两者具有不同的凋亡底物。因此，在死亡受体信号或其他的细胞过程中发挥各自的效应[21]。

除 caspase 的抑制剂，如痘病毒蛋白 CrmA 和杆状病毒蛋白 p35 特异作用于 caspase 外，哺乳细胞内源性 caspase 抑制剂 XIAP 也与 caspase3、7 相抵[22]。目前文献报道，caspase 至少由 3 种不同类型的抑制剂来调节。① FLIPs 含有无活性的 caspase 同源结构域，其催化位点已被破坏。这些蛋白作为显性失活的 Caspase 级联放大的调节蛋白而存在。② caspase 活性可以被 IAP（inhibitor of apoptosis proteins）所抑制，它包含 1～3 个不同的锌指结构，称为 BIR 结构域，可直接参与抑制凋亡。此外，它也包括一个 RING 指状结构，可能介导某些特异的蛋白-蛋白相互作用；某些 IAPs 还含 CARD 结构域，可能与 caspase 的 CARD 相互

作用。③ 另一个是近来发现的 ARC，仅含有 CARD 结构域[7]。

然而，有实验证实，细胞凋亡在缺乏 caspase 的状态下亦可发生。当加入 caspase 抑制剂时，细胞凋亡仍可发生[23]，这表明 caspase 激活以前，细胞已注定死亡。既然上游 caspase 向下游“执行者”传递信号是通过断裂酶原内结构域的连接子，由此推断，其他蛋白酶也可以通过相似的机制来激活凋亡途径。确实，细胞毒细胞蛋白酶颗粒酶 B 亦采用此机制。它是惟一的在哺乳细胞中作用于天冬氨酸的蛋白酶[24]。更为瞩目的是，没有天冬氨酸位点的蛋白酶在体外连接子区存在另一位点，也可激活 caspase 酶原。而 CD47 诱导 T 细胞凋亡的途径与 Fas、TNF 介导的信号无关，并不涉及 DNA 断裂和 caspase 激活[25]。因此，caspase 的凋亡调控机制仍需进一步研究。

3. 线粒体内 Bcl-2 家族蛋白的调节

线粒体过去简单地被认为是产生能量的细胞器，目前普遍认为是凋亡的枢纽，它通过释放死亡促进因子入胞质而调控着细胞的生死。由于线粒体以外膜与胞质相隔（compartment），因此，凋亡信号得以在不同区室间进行传导和易位（translocate），这种信号流的空间位移和延伸，在功能上使凋亡信号转导更加复杂化。

位于线粒体外膜上的 Bcl-2，与 ced-9 同源，是从滤泡 B 细胞瘤中克隆出的第一个拮抗凋亡的蛋白。随着凋亡研究的深入，Bcl-2 家族已经有许多新成员，包括拮抗凋亡的 Bcl-2、Bcl-xL 等和促进凋亡的 Bax、Bak 等蛋白[26]。多数 Bcl-2 家族成员在 C 端存在疏水结构域。至少，Bcl-2、Bcl-xL 的这一区域负责定位内质网、线粒体外膜和核外膜。此外，Bcl-2 家族的一个显著特征是易形成同源或异源二聚体[27]，促凋亡蛋白和抗凋亡蛋白的比率决定着细胞是否生存。例如：Bad 可与 Bcl-2、Bcl-xL 形成异二聚体，Bad 的过表达拮抗 Bcl-2、Bcl-xL 的生存作用；Bax 也可与 Bcl-2、Bcl-xL 形成异二聚体，尽管其亲和性较低[28]。目前，对这一类细胞生存调节蛋白的研究表明，它们控制线粒体的渗透转换（permeability transition，PT）和离子、蛋白的外流，如电子运输蛋白细胞色素 c[29]。细胞色素 c 一旦释放，可激活 caspase 家族。例如：Bax、Bak 刺激电压依赖的离子通道（VDAC）开放，细胞色素 c 可以穿过这一通道[30]。另有观点表明，由于线粒体外膜的破裂，通道的形成大得足以允许细胞色素 c 通过[31]。同时，Bcl-2 家族的蛋白联合不同的 ATPase（Ap-ATPase）共同调节凋亡的信号转导。Ap-ATPase 以复合物的方式激活 caspase，似乎涉及各自的 CARD 结构域间的相互作用，或 Ap-ATPase 结构域和蛋白酶结构域间的相互作用[7]。而 Bcl-2 家族蛋白 Diva、Boo 还可直接与 Apaf-1 相互作用[32]。总之，关键的调节蛋白是 Bcl-2 家族和凋亡 ATPase。目前，随着生物信息学的发展，通过查询 GenBank 数据库，已经鉴定出一个新的人 Bcl-2 样基因——Bcl2-L-10，它含有保守的 BH4、BH1 和 BH2 结构

域，但缺乏 BH3 结构域，定位在 15q21.2 染色体；通过阻止细胞色素 c 的释放和线粒体膜电位的降低来防止凋亡的产生，但不能阻滞 TNFα 诱导的凋亡；它的 BH4 结构域和跨膜结构域对其功能是必需的[33]。另有两个 BAG-1 蛋白家族成员 BAG-3 和 BAG-4，可以与 Bcl-2 结合。其中，BAG-4 与死亡结构域抑制蛋白 (suppressor of death domains, SODD) 相应，它可以抑制 TNFR1 的信号传导。BAG-3 和 BAG-4/SODD 都可以与 Bcl-2 发生物理性的相互作用[34]，其作用机制仍需进一步探讨。

介导细胞凋亡的 BH3 结构域的鉴定导致了 Bcl-2 家族的另一类成员的出现——即仅具有一个 BH3 同源结构域的蛋白。目前，哺乳动物至少有 7 个这样的蛋白，Bad、Bik、Blk、Hrk/DP5、Bid、Bim、Noxa，线虫的 EGL 仅含有 BH3 (9～16 个氨基酸) 结构域[35]。其中，线虫中 EGL 蛋白作为细胞凋亡的核心存在[36]。而缺陷 Bim 的小鼠甚至其杂合子，在发育 9.5d 前就死于宫腔内[37]。因此，仅含有 BH3 结构域的蛋白在哺乳动物的发育中起至关重要的作用。另有文献报道，Bad 和 Bim 的激活发生在凋亡早期，EGL-1 则作用于 ced-3 的上游。因而，仅含有 BH3 结构域的蛋白被认为是凋亡启动的基本组分，它在细胞内与进化保守的蛋白结合，整合细胞生存和死亡信号的调节网络，成为理想的调节不同细胞器和代谢过程的敏感因子[38]。

Bcl-G 是 Bcl-2 家族促凋亡蛋白中独特的新成员，其基因编码两个蛋白，长片段 Bcl-GL 和短片段 Bcl-Gs。两者在结构、表达、分布和功能上均存在差异。首先，Bcl-GL 含有 BH2、BH3 结构域（在 Bcl-2 家族中，没有仅含有 BH2、BH3 结构域而无 BH1 结构域的蛋白），它是惟一一个具有新结构特征的促凋亡蛋白。而 Bcl-Gs 则仅含有 BH3 结构域，属于仅含有 BH3 结构域的蛋白亚家族成员，很可能参与细胞凋亡的启动。其次，Bcl-GL 的 mRNA 可在几种正常组织中表达，而 Bcl-Gs 的 mRNA 仅在睾丸中表达。这也是 Bcl-G（Gonad）得以命名的原因，暗示 Bcl-Gs 的凋亡调节可能与性激素的信号转导通路相关。第三，Bcl-GL 弥散分布于胞质；尽管无膜定位的疏水区域，Bcl-Gs 依靠非 BH3 结构域定位在胞内的细胞器。这提示 Bcl-Gs 可能依靠与其他的蛋白相互作用来定位。最后，Bcl-G 的基因定位于染色体 12p12，此区域在 50% 的前列腺癌、30% 的卵巢癌和 30% 的儿童急性淋巴细胞白血病出现缺失，提示 Bcl-G 很可能还是肿瘤抑癌基因。当两者过表达时，Bcl-GL 所致的凋亡呈现中度的增加，而 Bcl-Gs 则显示很强的细胞毒性。缺失 BH2 结构域的 Bcl-GL 凋亡活性增强，并可与 Bcl-xL 发生免疫共沉淀，表明 BH2 结构域具有自我抑制功能。而依赖 BH3 结构域的 Bcl-Gs 可被共表达的抗凋亡蛋白 Bcl-xL 所抑制，且 Bcl-xL 也与 Bcl-Gs 发生免疫共沉淀，但对缺失 BH3 结构域的 Bcl-Gs 和 Bcl-GL 没有作用[39]。

某些 Bcl-2 家族蛋白是通过翻译后丝氨酸残基磷酸化来调节[28]，尽管其磷酸化的功能值得讨论。有几项研究表明，Bcl-2 通过磷酸化而失去活性，这似乎依

赖弹性环状结构域。弹性环状结构域由于丝氨酸的磷酸化，表现负向调节功能，反之，则增强 Bcl-2 的保护作用。Bad 的磷酸化也在丝氨酸残基，致使不能与 Bcl-xL 结合而发挥作用[40]。而另一些 Bcl-2 家族蛋白是通过断裂方式来调节。例如：caspase8 断裂 Bid，其 C 端易位至线粒体，在此促进细胞色素 c 的释放[41,42]。虽然对 Bcl-2 家族蛋白的调控机制研究很多，但是仍存在许多盲点，有待进一步研究。

4．细胞核内分子的参与

细胞核内某些基因也参与调节凋亡。不但 caspase3 直接断裂 RB 蛋白，而且，作为应激信号整合位点的 p53 可为 DNA 断裂和一些应激因素所激活。这表明细胞凋亡信号和细胞周期调控可能存在某种联系[43]。并且，p53 所触发的凋亡刺激 Bax 的表达，致使线粒体细胞色素 c 释放，一并激活 caspase9，与线粒体依赖的凋亡传导通路相通。除此之外，几种其他的转录因子（如：Myc 和 NF-κB 等）参与细胞死亡相连[7]。

c-Myc 作为细胞的转录因子已广为人知。近来，越来越多的文献报道其参与凋亡的调节。c-Myc 是 Myc 家族的主要成员，为三联体结构，在其末端结构域存在与细胞增殖、凋亡和转录活性相关的关键区域。其 C 端有一碱性/螺旋-环-螺旋（helix-loop-helix）/亮氨酸拉链（b/HLH/Z）基序，通过此区域介导寡聚化并识别所有靶基因上特定的 DNA 盒 CACGTGE。当它与 Max 形成异寡聚体时，才具有与 DNA 结合能力和生物活性。其 N 端包括 Myc 的转录激活结构域（transcriptional activation domain，TAD），含有两段 20 个氨基酸的区域，称为 Myc 盒 1 和 Myc 盒 2（Myc boxes 1 和 2、MB1 和 MB2），是 Myc 家族的保守区域，对生物活性具有关键作用。

c-Myc 可与众多凋亡相关蛋白发生联系。早在 1991 年，就有研究表明 c-Myc 的持续表达可以明显加速凋亡的产生[44]。其诱导的凋亡不但与一定的 JNK 激酶和 caspase3 有关，导致染色质凝聚和核 DNA 的断裂[45]，而且它还影响 caspase9，导致凋亡 DNA 内切酶 DFF/CAD 与抑制物 ICAD 形成复合物，最终被 caspase3 断裂和激活，但机制不明[46]。此外，c-Myc 可影响 TNFα 激活的凋亡和存活信号通路间的微妙平衡[47]。它可以阻碍 TRAF2 依赖的生存通路（负责 JNK 和 NF-κB 共激活），导致细胞对 TNFα 敏感，也可在 CD95/Fas 的上游和下游发挥作用[48]。Bcl-2 蛋白和 p53 也可通过影响细胞表达 c-Myc 来调控凋亡[49]。近来工作发现，c-Myc 与接头蛋白 Bin1 的 NTD 相互作用可以激活凋亡[50]。Bin1 在胞核和胞质间移动，以接头蛋白的形式将 c-Myc 与凋亡的死亡执行机制连接起来。然而，Bin1 在缺乏内切酶的肿瘤细胞中以 caspase 非依赖的机制诱导凋亡，与染色质凝聚和核 DNA 断裂无关。Pim-1 原癌基因编码丝/苏氨酸 Pim-1 蛋白，在肿瘤发生和凋亡过程中皆与 c-Myc 相关。有报道 Cdc25A 是 Pim-1 激酶的底

物[51]。仍需进一步的研究证明其促凋亡作用是如何被调控的。

5. 其他诱导凋亡的信号分子

几个丝/苏氨酸蛋白激酶也涉及凋亡。钙调蛋白依赖的死亡相关蛋白激酶(calmodulin-dependent death-associated protein kinase, DAP)包括一个DD和干扰素γ诱导的阳性调节子，调节增殖与凋亡间的平衡；另一个干扰素诱导的、双链RNA激活的丝/苏氨酸蛋白激酶PKR，在人类胚胎肾细胞中，对Fas介导的凋亡起作用[52]。此外，不但含有DD的RIP激酶在TRADD的下游与凋亡效应子RAIDD相互作用，激活另一凋亡通路，而且在RICK中，与CARD结构域相连的激酶结构域也涉及调节CD95/Fas和TNFR1依赖的凋亡途径[53]。而Rho家族GTP酶涉及细胞骨架的重构。Rho的灭活导致凋亡，它通过信号平衡协调细胞骨架的重构[54]。

神经酰胺作为对细胞凋亡反应的介导子引起广泛注意。由于各种应激和生物信号与其相关，具有复杂性和多样性[55]。一些学者认为，TNFR-55激活鞘磷脂酶，鞘磷脂酶酶解产生神经酰胺，导致其水平增高[56]，发生凋亡。在TCR信号转导中，神经酰胺的产生亦发挥重要作用[57]。与之相反，另一些学者认为，神经酰胺作为应激反应途径的转导子存在，而凋亡的发生仅是一个潜在的结果[58]。例如：在U937细胞中，研究神经酰胺和TNFα的关系中发现，神经酰胺在TNFα介导的凋亡中，不是第二信使[59]。同时，利用高分辨率质谱分析测试神经酰胺，表明它作为凋亡事件的结果出现[60]。

另有一些凋亡相关蛋白被发现。例如：在星型胶质细胞中，cAMP依赖的信号转导通路调节螺旋-环-螺旋转录因子Id4，其过表达可诱导凋亡[61]。而在前列腺癌的治疗研究中，Par-4（prostate apoptosis response-4）可以选择性地诱导PC-3、DU-145和TSU-Pr细胞发生凋亡，初步研究表明，Par-4通过抑制NF-κB的活性及激活Fas诱导凋亡的信号传导通路，导致肿瘤的抑制[62]。

二、与细胞内其他信号转导通路的相互作用

大量文献报道，细胞凋亡的信号转导与细胞增殖、分化等其他的重要生命活动相关，如MAPK、JAK/STAT、JNK、蛋白激酶B、C等通路。这种多信号分子的参与、多同工型部分的重叠、各信号分子间的空间转换导致了凋亡信号转导的复杂性。

一方面，ALP (alkyl-lysophospholipid)、HGF和TNFR2通过激活促凋亡的SAPK/JNK、JNK途径导致凋亡的增加[63-65]，而在VEGF调节凋亡途径中，通过激活ERK1、ERK2，抑制SAPK/JNK来决定内皮细胞是否凋亡[66]。另外，不仅在乳腺癌细胞中，抗癌药物POH通过诱导TGFα和Smad3参与凋亡信号传

导，导致肿瘤转归[67]，PKC 还可以通过调节 FADD 的募集，改变 DISC 的形成效率而调节 Fas/CD95 所诱导的凋亡信号[68]。

另一方面，在缺氧的条件下，端粒酶的激活伴随 MAPK 的激活和 c-fos 的表达来拮抗凋亡[69]。并且，蛋白激酶 B（c-Akt）、PI-3 激酶、STAT5 的某些组分可以被激活来拮抗凋亡[70]，例如：IGF、黏着斑激酶（focal adhesion kinase，FAK）和一种质膜肌动蛋白细胞骨架连接蛋白 Ezrin，通过激活其下游的 Akt 激酶来抑制凋亡[71～73]。不但在 NO 诱导的凋亡研究中，PKC 同工型的过表达阻止凋亡，这涉及 JNK/SAPK、p38 激酶和 cpp32 蛋白酶途径[74]，而且在人类嗜酸性粒细胞中，GM-CSF 拮抗凋亡，其中 JAK2 的激活起关键作用[75]。

细胞发育、应激和其他各种诱导刺激的信号，经过各种转导子，在细胞膜和细胞核间发生空间的转变；信号从最初微量刺激到蛋白酶级联放大的量变；细胞内各种转导子间及与细胞分化、增殖分子之间相互的反馈调节；从凋亡的触发到凋亡形成发生的质变，其信号转导网络的复杂性可见一斑。然而，对于各种转导途径如何交叉对话（cross-talk），对于触发、调节网络以及其中心机制，对于凋亡信号转导的基因调控尚存在诸多疑问，仍需进一步深入研究，揭示其内在本质。

参 考 文 献

1 Kerr J F, Wyllie A H, Currie A R. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. Br J Cancer, 1972, 26: 239～254

2 Nicholson D W. ICE/CED3-like proteases as therapeutic targets for the control of inappropriate apoptosis. Nat Biotechnol, 1996, 14 (3): 297～301

3 Prasad S C, Soldatenkov V A, Kuettel M R, Thraves P J, Zou X, Dritschilo A. Protein changes associated with ionizing radiation-induced apoptosis in human prostate epithelial tumor cells. Electrophoresis, 1999, 20: 1065～1074

4 Gupta S. Molecular steps of death receptor and mitochondrial pathways of apoptosis. Life Sci, 2001, 69 (25-26): 2957～2964

5 Weng G Z, Bhalla U S, Lyengar R. Complexity in biological signaling systems. Science, 1999, 284 (2): 92～95

6 Kumar S, Colussi P A. Prodomains-adaptors-oligomerization: the pursuit of caspase activation in apoptosis. TIBS, 1999, 24: 1～4

7 Aravind L, Dixit V M, Koonin E V. The domains of death: evolution of the apoptosis machinery. TIBS, 1999, 24: 47～53

8 Weber C H, Vincenz C. The death domain superfamily: a tale of two interfaces? Trends Biochem Sci, 2001, 26 (8): 475～481

9 Tartaglia L A, Ayres T M, Wong G H, Goeddel D V. A novel domain within the 55 kd TNF receptor signals cell death. Cell, 1993, 74: 845～853

10 Baker S J, Reddy E P. Modulation of life and death by the TNF receptor superfamily. Oncogene, 1998, 17: 3261～3270

11 Hsu H, Xiong J, Goeddel D V. The TNF receptor 1-associated protein TRADD signals cell death and NF-kappa B activation. Cell, 1995, 81: 495～504

12 Chinnaiyan A M, O' Rourke K, Tewari M, Dixit V M. FADD, a novel death domain-containing protein, interacts with the death domain of Fas and initiates apoptosis. Cell, 1995, 81: 505～512

13 Yang X, Khosravi-Far R, Chang H Y, Baltimore D. Daxx, a novel Fas-binding protein that activates JNK and apoptosis. Cell, 1997, 89: 1067～1076

14 Stanger B Z, Leder P, Lee T H, Kim E, Seed B. RIP: a novel protein containing a death domain that interacts with Fas/APO-1 (CD95) in yeast and causes cell death. Cell, 1995, 81: 513～523

15 Thome M, Hofmann K, Burns K, Martinon F, Bodmer J L, Mattmann C, Tschopp J. Identification of CARDIAK, a RIP like kinase that associates with caspase-1. Cur Biol, 1998, 16: 885～888

16 Jung Y S, Kim K S, Kim K D, Lim J S, Kim J W, Kim E. Apoptosis-linked gene 2 binds to the death domain of Fas and dissociates from Fas during Fas-mediated apoptosis in Jurkat cells. Biochem Biophys Res Commun, 2001, 288 (2): 420～426

17 Thornberry N A, Lazebnik Y. Caspases: enemies within, Science, 1998, 281: 1312～1316

18 Chou J J, Matsuo H, Duan H, Wagner G. Solution structure of the RAIDD CARD and model for CARD/CARD interaction in caspase-2 and caspase-9 recruitment. cell, 1998, 94: 171～180

19 Salvesen G S, Dixit V M. Caspases: intracellular signaling by proteolysis. Cell, 1997, 91 (14): 443～446

20 Woo M, Hakem R, Soengas M S, Duncan G S, Shahinian A, Kagi D, Hakem A, McCurrach M, Khoo W, Kaufman S A, Senaldi G, Howard T, Lowe S W, Mak T W. Essential contribution of caspase 3/CPP32 to apoptosis and its associated nuclear changes. Gene Dev, 1998, 12: 806～819

21 Wang J, Chun H J, Wong W, Spencer D M, Lenardo M J. Caspase-10 is an initiator caspase in death receptor signaling. Proc Natl Acad Sci USA, 2001, 98 (24): 13884～13888

22 Deveraux Q L, Takahashi R, Salvesen G S, Reed J C. X-linked IAP is a direct inhibitor of cell-death proteases. Nature, 1997, 388: 300～304

23 Denmeade S R, Lin X S, Tombal B, Isaacs J T. Inhibition of caspase activity does not prevent the signaling phase of apoptosis in prostate cancer cells. Prostate, 1999, 39 (4): 269～279

24 Zhivotovsky B, Burgess D H, Vanags D M, Orrenius S. Involvement of cellular proteolytic machinery in apoptosis. Biochem Biophys Res. Commun, 1997, 230: 481～488

25 Pettersen R D, Hestdal K, Olafsen M K, Lie S O, Lindberg F P. CD47 signals T cell death. J Immunol, 1999, 162 (12): 7031～7040

26 Adams J M, Cory S. The Bcl-2 protein family: arbiters of cell survival. Science, 1998, 281: 1322～1326

27 Gajewski T F, Thompson C B. Apoptosis meets signal transduction: elimination of a BAD influence. Cell, 1996, 87: 589～592

28 Kiefer M C, Brauer M J, Powers V C, Wu J J, Umansky S R, Tomei L D, Barr P J. Modulation of apoptosis by the widely distributed Bcl-2 homologue Bak. Nature, 1995, 374: 736～739

29 Chao D T, Korsmeyer S J. Bcl-2 family: regulators of cell death. Ann Rev Immunol, 1998, 16: 395～419

30 Shimizu S, Narita M, Tsujimoto Y. Bcl-2 family proteins regulate the release of apoptogenic cytochrome c by the mitochondrial channel VDAC. Nature, 1999, 399: 483～487

31 Martinou J C. Key to the mitochondrial gate. Nature, 1999, 399: 411-412

32 Inohara N, Gourley T S, Carrio R, Muniz M, Merino J, Garcia I, Koseki T, Hu Y, Chen S, Nunez G. Diva, a Bcl-2 homologue that binds directly to Apaf-1 and induces BH3-independent cell death. J Biol Chem, 1998, 273: 32479～32486

33 Zhang H, Holzgreve W, De Geyter C. Bcl2-L-10, a novel anti-apoptotic member of the Bcl-2 family, blocks

apoptosis in the mitochondria death pathway but not in the death receptor pathway. Hum Mol Genet, 2001, 10 (21): 2329～2339

34 Antoku K, Maser R S, Scully W J Jr, Delach S M, Johnson D E. Isolation of Bcl-2 binding proteins that exhibit homology with BAG-1 and suppressor of death domains protein. Biochem Biophys Res Commun, 2001, 286 (5): 1003～1010

35 Lutz R J. Role of the BH3 (Bcl-2 homology 3) domain in the regulation of apoptosis and Bcl-2 related proteins. Biochem Socio Trans, 2000, 28 (2): 51～56

36 Conradt B, Horvitz H R. The C. elegans protein EGL-1 is required for programmed cell death and interacts with the Bcl-2-like protein CED-9. Cell, 1998, 93 (4): 519～522

37 Bouillet P, Metcalf D, Huang D C, Tarlinton D M, Kay T W, Kontgen F, Adams J M, Strasser A. Proapoptotic Bcl-2 relative Bim required for certain apoptotic responses, leukocyte homeostasis, and to preclude autoimmunity. Science, 1999, 286 (5445): 1735-1738

38 David C, Huang S, Strasser A. BH3-only proteins—essential initiators of apoprotic cell death. Cell, 2000, 103: 839～842

39 Guo B, Godzik A, Reed J C. Bcl-G, a novel pro-apoptotic member of the Bcl-2 family. J Biol Chem, 2001, 276 (4): 2780～2785

40 Muchmore S W, Sattler M, Liang H, Meadows R P, Harlan J E, Yoon H S, Nettesheim D, Chang B S, Thompson C B, Wong S L, Ng S L, Fesik S W. X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. Nature, 1996, 381: 335～341

41 Luo X, Budihardjo I, Zou H, Slaughter C, Wang X. Bid, a Bcl-2 interacting protein, mediates cytochrome c release from mitochondria in response to activation of cell surface death receptors. Cell, 1998, 94 (21): 481-490

42 Chou J J, Li H, Salvesen G S, Wagner G. Solution structure of Bid, an intracellular amplifier of apoptotic signaling. Cell, 1999, 96 (5): 615～624

43 Meek D W. New developments in the multi-site phosphorylation and integration of stress signaling at p53. Int J Radia Biol, 1998, 74 (6): 729～737

44 Askew D S, Ashmun R A, Simmons B C, Cleveland J L. Constitutive c-Myc expression in an IL-3-dependent myeloid cell line suppresses cell cycle arrest and accelerates apoptosis. Oncogene, 1991, 6: 1915～1922

45 Yu K, Ravera C P, Chen Y N, McMahon G. Regulation of Myc-dependent apoptosis by p53, c-Jun N-terminal kinases/stress-activated protein kinases, and Mdm-2. Cell Growth Differ, 1997, 8 (7): 731～742

46 Enari M, Sakahira H, Yokoyama H, Okawa K, Iwamatsu A, Nagata S. A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. Nature, 1998, 391: 43～50

47 Klefstrom J, Arighi E, Littlewood T, Jaattela M, Saksela E, Evan G I, Alitalo K. Induction of TNF-sensitive cellular phenotype by c-Myc involves p53 and impaired NF-kappa B activation. EMBO J, 1997, 15: 7382～7392

48 Ashkenazi A, Dixit V M. Death receptors: signaling and modulation. Science, 1998, 281: 1305～1308

49 Zindy F, Eischen C M, Randle D H, Kamijo T, Cleveland J L, Sherr C J, Roussel M F. Myc signaling via the ARF tumor suppressor regulates p53-dependent apoptosis and immortalization. Genes Dev, 1998, 12: 2424～2433

50 Prendergast G C. Mechanisms of apoptosis by c-Myc. Oncogene, 1999, 18: 2967～2987

51 Mochizuki T, Kitanaka C, Noguchi K, Muramatsu T, Asai A, Kuchino Y. Physical and functional interactions between Pim-1 kinase and Cdc25A phosphatase. Implications for the Pim-1-mediated activation of the c-Myc signaling pathway. J Biol Chem, 1999, 274 (26): 18659～18666

52 Takizawa T, Tatematsu C, Nakanishi Y. Double-stranded RNA-activated protein kinase (PKR) fused to green fluorescent protein induces apoptosis of human embryonic kidney cells: possible role in the Fas signaling pathway. J Biochem, 1999, 125 (2): 391～398

53 Deiss L P, Feinstein E, Berissi H, Cohen O, Kimchi A. Identification of a novel serine/threonine kinase and a novel 15-kD protein as potential mediators of the gamma interferon-induced cell death. Gene Dev, 1995, 9: 15～30

54 Moorman J P, Luu D, Wickham J, Bobak D A, Hahn C S. A balance of signaling by Rho family small GTPases RhoA, Rac1 and Cdc42 coordinates cytoskeletal morphology but not cell survival. Oncogene, 1999, 18 (1): 47～57A

55 Okazaki T, Kondo T, Kitano T, Tashima M. Diversity and complexicity of ceramide signaling in apoptosis. Cell Signal, 1998, 10 (10): 685-692

56 Hofmann K, Dixit V M. Ceramide in apoptosis—does it really matter? TIBS, 1998, 23: 374～377

57 Tonnetti L, Veri M C, Bonvini E, D' Adamio L. A role for neutral sphingomyelinase-mediated ceamide production in T cell receptor-induced apoptosis and mitogen activated protease kinase mediated signal transduction. J Exp Med, 1999, 189 (10): 1581～1589

58 Kolesnick R, Hannun Y A. Ceramide and apoptosis. TIBS, 1999, 24: 224～225

59 Karasavvas N, Zakeri Z. Relationships of apoptotic signaling mediated by ceramide and TNFa in U937 cells. Cell Death Differ, 1999, 6 (2): 115～123

60 Watts J D, Gu M, Patterson S D, Aebersold R, Polverino A J. On the complexities of ceramide changes in cells undergoing apoptosis: lack of evidence for a second messenger function in apoptotic induction. Cell Death Differ, 1999, 6 (2): 105～114

61 Andress-Barquin P J, Hernandez M C, Israel M A. Id4 expression induces apoptosis in astrocytic cultures and is down-regulated by activation of the cAMP-dependent signal transduction pathway. Exp Cell Res, 1999, 247 (2): 347～355

62 Chakraborty M, Qiu S G, Vasudevan K M, Rangnekar V M. Par-4 drives trafficking and activation of Fas and Fasl to induce prostate cancer cell apoptosis and tumor regression. Cancer Res, 2001, 61 (19): 7255～7263

63 Ruiter G A, Zerp S F, Bartelink H, van Blitterswijk W J, Verheij M. Alkyl-lysophospholipids activate the SAPK/JNK pathway and enhance radiation-induced apoptosis. Cancer Res, 1999, 59 (10): 2457～2463

64 Conner E A, Teramoto T, Wirth P J, Kiss A, Garfield S, Thorgeirsson S S. HGF-mediated apoptosis via p53/bax-independent pathway activating JNK1. Carcinogenesis, 1999, 20 (4): 583～590

65 Jupp O J, McFarlane S M, Anderson H M, Littlejohn A F, Mohamed A A, MacKay R H, Vandenabeele P, MacEwan D J. Type II tumor necrosis factor-alpha receptor (TNFR2) activates c-Jun N-terminal kinase (JNK) but not mitogen-activated protein kinase (MAPK) or p38 MAPK pathways. Biochem J, 2001, 359 (Pt 3): 525～535

66 Gupta K, Kshirsagar S, Li W, Gui L, Ramakrishnan S, Gupta P, Law P Y, Hebbel R P. VEGF prevents apoptosis of human microvascular endothelial cells via opposing effects on MAPK/ERK and SAPK/JNK signaling. Exp Cell Res, 1999, 247 (2): 495～504

67 Ariazi E A, Satomi Y, Ellis M J, Haag J D, Shi W, Sattler C A, Gould M N. Activation of the transforming growth factor beta signaling pathway and induction of cytostassi and apoptosis in mammary carcinomas treated with the anticancer agent perillyl alcohol. Cancer Res, 1999, 59 (8): 1917～1928

68 Gomez-Angelats M, Cidlowski J A. Protein kinase C regulates FADD recruitment and death-inducing signaling complex formation in Fas/CD95-induced apoptosis. J Biol Chem, 2001, 276 (48): 44944～44952

69 Seimiya H, Tanji M, Oh-hara T, Tomida A, Naasani I, Tsuruo T. Hypoxia up-regulates telomerase activity via mitogen-activated protein kinase signaling in human solid tumor cells. Biochem Biophycs Res Commun, 1999, 260 (2): 365～370

70 Bao H, Jacobs-Helber S M, Lawson A E, Penta K, Wickrema A, Sawyer S T. Protein kinase B (c-Akt), Phosphatidylinositol 3-kinase, and STAt5 are activated by erythropoietin in HCD57 erythroid cells but are constitutively active in an EPO-independent, apotosis-resistant subclone (HCD57-SREI cells). Blood, 1999, 93 (11): 3757～3773

71 Rohn J L, Hueber A O, McCarthy N J, Lyon D, Navarro P, Burgering B M, Evan G I. The opposing roles of the Akt and c-Myc signaling pathways in survival from CD95-mediated apoptosis. Oncogene, 1998, 17 (22): 2811～2818

72 Gautreau A, Poullet P, Louvard D, Arpin M. Ezrin, a plasma membrane-microfilament linker, signals cell survival through the phosphatiylinositol 3-kinase/Akt pathway. Proc Natl Acad Sci USA, 1999, 96 (13): 7300～7305

73 Cary L A, Guan J L. Focal adhesion kinase in integrin-mediated signaling. Front Biosci, 1999, 15 (4) D102～113

74 Guo Y L, Kang B, Williamson J R. Resistance to TNF-a cytotoxicity can be achieved through different signaling pathways in rat cells. Am J Physiol, 1999, 1423 (1): R29～33

75 Catlett-Falcone R, Landowski T H, Oshiro M M, Turkson J, Levitzki A, Savino R, Ciliberto G, Moscinski L, Fernandez-Luna J L, Nunez G, Dalton W S, Jove R. Constitutive activation of Stat3 signaling confers resistance to apoptosis in human U266 myeloma cells. Immunity, 1999, 10 (1): 105～115

第二节　细胞凋亡的蛋白质组研究

尽管 mRNA 差异显示（differential display）技术有效，但仍存在着假阳性和差别条带小的缺限[1]。一些新的技术，如基因表达连续分析法（serial analysis of gene expression, SAGE）和 DNA 微阵列分析法（DNA microarray analysis）相继被用来阐明基因组的功能。然而，上述研究属于功能基因组计划研究领域，不能涵盖① 基因的表达产物 mRNA 的贮存、转运、降解、翻译调控；② 基因的最终产物/基因功能的执行体——蛋白质的化学修饰（如糖基化、磷酸化）、剪切加工（如酶原降解、结构域拼接）；③ 蛋白质间相互作用等诸多领域。而生命活动的真正执行体是蛋白质，与基因组学相比，蛋白质组学研究更接近生命的本来面目。

蛋白质组（proteome），是澳大利亚科学家 Marc Wilkins[2] 于 1995 年选择“PROTEin”与“genOME”这两个单词的一部分创造出的新概念，它随着组织、细胞及环境状态的不同而改变，可以反映生命活动某一时刻的瞬时表达状态，它的研究已涉及生命科学中的一系列热点领域，使人们对生命活动的机制认识由间接的基因、核酸层次深入到蛋白质层次[3]。目前，双向电泳是蛋白质组研究中惟一的蛋白分离技术，IPG 的出现使双向电泳的重复性和分辨率大大提高。传统的 2-DE 在鉴定蛋白的时候要求同时制备 10 块凝胶，并收集足够的蛋白量，既

耗时又容易出现斑点变异。目前，多采用增加样品量和缩窄 pH 范围，可以满足质谱鉴定的要求[4]。分离后的蛋白质鉴定是利用 MALDI-TOF、肽质量指纹谱技术（peptide mass fingerprinting，PMF）在 fmol 水平对肽片段混合物进行检测，并联合数据库查寻来鉴定蛋白。然而，这种方法仅成功适用存在于核酸和蛋白质数据库内的已知蛋白；而且，某些肽片段的修饰使这种鉴定的可信性降低。由于 PMF 的局限性，应联合微量测序、氨基酸组分分析和 Western Blot 等进行鉴定。一旦鉴定出某一领域内的具有特征性的候选蛋白，即为功能提供线索，并对今后深入研究蛋白的功能奠定基础。在短短的几年间，这一领域的研究工作进展迅速，已分别从流感嗜血杆菌（*Haemophilus influenzae*）、大肠杆菌（*Escherichia coli*）、线虫（*Caenorhabditis elegans*）和果蝇（*Drosophila melanogaster*）检出 300 个、1600 个、2000 个和 1200 个蛋白质斑点，其中大部分蛋白质得到鉴定[5]。由于这些技术的进步，使获得高重复性的蛋白表达谱和进行适度的生物信息学分析成为可能，现在它已成为一种非常有效的蛋白质差异显示及分析技术。

近年来，蛋白质组技术也渐渐被引入到细胞凋亡的领域，用于筛选与凋亡相关的蛋白。在研究的初始，仅将双向电泳技术用于分离各类细胞凋亡模型中的蛋白。在人神经胶质瘤细胞和鸡睫状神经节神经元中撤除肥大因子所诱导的凋亡中，分离出部分与凋亡相关的蛋白[6,7]。对出生后 1 天的 Wistar 大鼠的颈上交感神经节神经元撤除神经生长因子后所导致的凋亡，双向电泳可分离出 17 个表达降低和 4 个表达增加的蛋白，但未作进一步的鉴定[8]。而在 TNF 联合蛋白合成抑制剂诱导小牛主动脉内皮细胞凋亡、cAMP 诱导大鼠早幼粒白血病细胞 IPC-81 凋亡中，应用双向电泳分离蛋白的同时，联合磷酸化酶抑制剂，提示蛋白的去磷酸化可能伴随凋亡[9]。另一方面，也应用双向电泳技术分离特定蛋白的同工型组分。在电离辐射诱导雌激素受体阴性的 MDA-MB-468 乳腺癌细胞的凋亡研究中，利用双向电泳联合抗体直接作用于 PARP 来探测其断裂位点，在 N 末端分离出至少由 3 个同工酶组成的相对分子质量为 24kDa 的片段[10]。利用牛磺酸或地塞米松诱导人 Jurkat 细胞凋亡，Bak 蛋白隐藏的 N 末端抗原决定簇暴露出来，表明 Bcl-xL 与 Bak 蛋白的解离影响其抗原的暴露[11]。人们并不局限于双向电泳分离的初级表象，理所当然地进一步联合各种质谱技术或 Western Blot 技术来进行分离后鉴定。在 Burkitt 淋巴瘤细胞中，联合纳喷串联质谱鉴定出与凋亡相关的蛋白：糖代谢相关酶、细胞骨架蛋白、分子伴侣、以及核糖体蛋白的某些组分等[12]。而在电离辐射诱导永生化的人前列腺上皮细胞 267B1-XR 凋亡中，则联合 Western Blot 技术，在凋亡执行中鉴定出中等纤维蛋白角蛋白 18、19、波形蛋白和相关的 14-3-3 接头蛋白水平降低，而位于内质网的分子伴侣如葡萄糖调节蛋白 94、肌钙网蛋白、钙连接蛋白和二硫异构酶表达增加[4]。虽然早期的细胞凋亡蛋白质组工作研究表浅，但为现今工作的研究奠定了基础。

随着蛋白质组技术的完善与发展，细胞凋亡的蛋白质组研究在质和量上都较

前人的研究有所超越。以抗 IgM 敏感的人 Burkitt 淋巴细胞系 BL-60 作为研究对象，用抗 IgM 处理细胞后，通过高流量的双向电泳进行分离比较。凋亡与非凋亡细胞的表达差异约有 80 个蛋白斑点。质谱和 Edman 测序鉴定表明，D4-GDI，一种含量丰富的、与 Ras 相关的 Rho 家族 GTPase 的造血干细胞 GDP 解聚抑制剂在加入抗 IgM 之后消失，迅速断裂成 23 kDa 的片段[13]。在 cAMP 诱导的大鼠 IPC-81 急早幼粒白血病细胞凋亡的研究中，联合高灵敏度的同位素[^{35}S]标记的甲硫氨酸检测双向电泳的分离，结果显示两个蛋白发生了翻译后修饰[14]。目前，蛋白质组研究已从简单的两两比较渐渐深入到多层次的比较，期望得到更丰富的信息。例如：利用双向电泳技术，研究成人的 7 个不同脑区内热激蛋白的表达谱，通过 MALDI-TOF 鉴定了 120 个蛋白。其中热激蛋白 70RY、同族热激蛋白（HSC）71 和葡萄糖调节蛋白（GRP）75 在多发性硬化（DS）的颞叶皮质区域表现出显著性降低（$P<0.05$），而 HSP70.1 和 GRP78 在小脑区域则表现为显著性增加（$P<0.05$）。当 TCP-1 ε 亚单位在顶叶皮质区表现为显著性降低（$P<0.05$）时，GRP78 在小脑区呈现相似程度的增加（$P<0.05$）。另外，α-crystallin B、HSP60 和 GRP94 并未表现出明显的改变。这篇文章首次在蛋白水平对各脑区进行观察，有 9 个不同的分子伴侣同时发生了改变，这为进一步研究多发性硬化提供了证据[15]。另一个研究小组与其具有异曲同工之妙。通过双向电泳技术，对 p53 基因剔除小鼠（p53-/-）与 p53 野生型小鼠（p53+/+）的 6 种类型的脑组织（颞叶皮质、小脑、海马、纹状体、嗅球和颈髓）、肺和血液进行比较和分析，结果显示，7/886 个蛋白斑点特异表达于 p53-/-小鼠的脑组织中，而 12/1008 个蛋白斑点特异表达于肺组织中。N 端氨基酸测序结果显示，在所有 p53-/- 小鼠中，一种新鉴定的小鼠线粒体 NADH-泛素氧化还原酶 24 kDa 亚单位表达下降，脂蛋白 A1 酸性形式表达增加。此外，脑型肌酸激酶 B 链和微管蛋白 β-5 N 端片段在 p53-/-小鼠的小脑区域增加，并且一个新蛋白 hydroxyacyl glutathione hydrolase（glyoxalase II）在 p53-/-小鼠的颞页皮质区表达降低。这些蛋白的变化表明与 p53 相关的功能，为进一步分析 p53 在中枢神经系统的功能提供新思路[16]。在进行比较细胞总蛋白的蛋白质组研究之外，有些实验室关注亚细胞器的分级分离。在研究渗透转换核孔复合物（PTPC）的开放过程中，已有 79 个已知蛋白被鉴定出来，此外，另有 21 个蛋白与 EST 相匹配。有几个蛋白直接或间接参与促凋亡过程，例如，endozepine（可促进线粒体膜通透性）、X-linked deafness dystonia protein 和葡萄糖调节蛋白 75（参与蛋白运输）、以及许多参与代谢的酶。尽管这些蛋白对凋亡的影响仍不明确，但是目前从 PTPC 释放出来的线粒体蛋白数据库应有助于阐明凋亡过程[17]。而在 Fas 诱导的 Jurkat T 淋巴细胞凋亡的蛋白质组研究中，利用双向电泳技术分离出 1000 个蛋白，图像分析比较对照组与凋亡组，其中 19 个蛋白的信号强度降低甚至消失，并新出现了 38 个蛋白斑点。进而从蛋白质从头合成、磷酸化状态、细胞内定位和亚细胞组分分级

等几个方面作进一步分析。结果表明，热激蛋白 27、70B、钙调蛋白和 H-Ras 合成蛋白依赖于 Fas 的诱导；34 个蛋白受到去磷酸化的影响，如：endoplasmin 和磷酸化、热激蛋白 70、57 和 90；位于细胞核的 Annexin IV 移位至细胞质，而 TCP-1α 在细胞质的信号强度降低，并可在细胞核内检测到。此外，在 12 个降解蛋白中，发现肌球蛋白重链作为 caspase 的新靶点，与其他诱导的细胞凋亡相比具有特定的改变。在这些研究中，研究者提出这样一个观点：发生凋亡的细胞是否都经历程序性凋亡，还是另辟蹊径[18]。与上述多角度研究相媲美的是另有一个研究小组用生物信息学分析方法整合数据。首先，利用消减双向电泳技术比较发生凋亡和对照组之间的蛋白改变，有 45 个蛋白发生差异表达。经过 MALDI-TOF 鉴定，结果表明，hnRNPs A0、A2/B1、A3、K 和 R，剪接因子 p54（nrb）、SRp30c、ASF-2 和 KH-type 剪接调节蛋白（FUSE-binding protein 2）参与凋亡；而 α-NAC、NS1-相关蛋 1、poly（A）-结合蛋白 4 至今并未涉及凋亡。进一步通过生物信息学分析，查寻 pfam 数据库显示，已鉴定的 15/21 个蛋白含有 RNP 或 KH motif，具有 RNA-结合蛋白的基本单位，而 12 个蛋白涉及剪接过程。这篇文章通过亚细胞分级、分离联合蛋白质组技术，有效地鉴定了凋亡相关的蛋白，并且通过生物信息学的分析，进一步提示凋亡相关蛋白最明显的特征是 RNA-binding 基序的出现。在这 12/15 个蛋白中，很可能具有 RNA 的剪接功能，涉及 RNA 的修饰和稳定[19]。因此，从所获得的各层面数据来看，在凋亡过程中所发生的广谱改变可能证实了一些关于细胞凋亡信号转导的生物学猜想。

此外，在其他领域的蛋白质组研究中，也能鉴定出与凋亡相关的蛋白。哺乳动物的线粒体核糖体（mitoribosome）富含 55S 系数蛋白组分，由 39S 和 28S 大小亚基组成。在鉴定 28S 小亚基蛋白的研究中，首先通过高度还原的自由基双向电泳分离，随后通过液相色谱/串联质谱，用电喷雾离子阱质谱及联合 EST 数据库搜寻来鉴定其 cDNA 序列。已经鉴定了 21 个蛋白，其中 11 个新蛋白来源于人和小鼠。其中，与凋亡相关的 DAP3 被鉴定出来，表明 mitoribosome 可能参与凋亡[20]。利用蛋白质组学的方法，建立了树突状细胞分泌 exosome 的蛋白表达谱，已有 21 个蛋白鉴定出来。其中，有一类蛋白与凋亡相关，包括硫氧还蛋白过氧化物 II、Alix、14-3-3 和 galectin-3。这些结果提示了 exosome 和凋亡细胞所释放的微颗粒之间存在某种关联[21]。利用蛋白质组方法还研究了 PP（peroxisome proliferator）处理的野生型和 PPARα（peroxisome proliferator activated receptor α）的肝细胞。在已鉴定的 18 个差异蛋白斑点中，抗凋亡蛋白葡萄糖调节蛋白 94（GRP94）发生了过表达。这提示存在通过 PP 激活抗凋亡机制的分子途径[22]。

虽然已经对各种细胞凋亡进行了广泛深入的研究，但是关于触发凋亡机制的中心问题之一——促使凋亡发生不可逆生化事件的启动机制和各种信号是如何整合并启动凋亡却知之甚少。近来报道，含有 BH3 结构域的 Bcl-2 家族蛋白相互作用，可能导致拓扑结构的变化，控制不可逆凋亡发生于凋亡执行时相之前，并对

细胞设置了一个生存的界域，表明它与凋亡过程本身的分离[11]。除此之外，细胞内是否发生了某些其他变化而导致不可逆凋亡，是否可以作用于凋亡启动靶点而在肿瘤凋亡治疗中发挥独特的作用至今尚不得而知。本实验室以寻找和研究细胞凋亡启动中的调控蛋白为目的，以全反式维甲酸（all-trans retinoic acid，ATRA）诱导急早幼粒白血病细胞 HL-60 为研究对象，利用蛋白质组技术，筛选与之相关的凋亡启动蛋白。

全反式维甲酸由维生素 A 衍生而来，是诱导分化剂中最重要并已用于临床治疗的一类药物。在临床上，ATRA 用于治疗急早幼粒白血病（acute promyelocytic leukemia，APL），患者的缓解率可达到 90%。超微结构观察发现，向粒系分化，产生核分叶的现象，并形成广泛膜限制的染色质片层（envelope-limited chromatin sheets，ELCS），表现为单层紧密包裹的 30 nm 的染色质纤维与核膜内外膜相连[23]。即首先发生分化，随后以凋亡的形式死亡[24]。目前，ATRA 作用于细胞的分子机制表明：ATRA 激活配体依赖的转录因子维甲酸受体（RAR），RAR 与 RXR 形成同源或异源二聚体，识别维甲酸受体反应元件（RARE）核心序列，通常是启动子的 RARE 区域，包括两个相同的六核苷酸“半位点”或识别序列（PuGGTCA）同向重复，其中间隔 1、2 或 5 个核苷酸。因此依靠配体结合，核受体识别和结合 RARE，激活或抑制基因表达[25]。既然 ATRA 对 RAR、RXR 传导通路进行激活，可以推测，一大群受两者调控的基因将会被抑制或诱导。而这些基因中很有可能包含有对于分化和凋亡至关重要的开关基因。近来研究表明各种已知的凋亡相关蛋白与 ATRA 所诱导的凋亡均存在某种联系。首先，在细胞终末分化朝向成熟粒系分化的过程中，抗凋亡蛋白 Bcl-2 和 Bcl-x 逐渐下调，甚至停止表达，而死亡促进蛋白 Bax 却存在[26]。此外，应用特异性的 caspase 抑制剂 DEVD-CHO 和 YVAD-CMK 进行研究，发现可以抑制由 ATRA 诱导的自发性凋亡[27]。而在 ATRA 诱导大鼠 S-V 细胞凋亡中，caspase3 活性增加[11]。其次，负责增殖调控的 Ras/Raf/MAPK 与其细胞凋亡相关的报道日渐增多。ATRA 通过 RAR 和 RXR 来抑制 v-src，调节 MAPK 和 ERK，来诱导基质金属蛋白酶-1（MMP-1）[28]。它还可以在各种细胞类型中诱导组织转氨酶的表达，在转录水平调节，并需要 RAR、RXR 配体依赖的信号调节[29]。并且，ATRA 诱导 T 淋巴细胞白血病（ATL）的研究不但涉及 p53 水平降低[30]，而且在髓性白血病细胞中通过抑制 IL-6 及其受体的表达、增加 Fas 的表达来诱导凋亡[31]，甚至细胞的还原状态及其氧化和抗氧化之间的平衡也明显影响 ATRA 对细胞的作用[32]。因此，ATRA 诱导 HL-60 细胞凋亡启动为促凋亡和抗凋亡级联“瀑布式”调节的信息调控提供了一个比较方便的研究模型[16]。

在上述研究的基础上，本实验室首先建立了 ATRA 诱导 HL-60 细胞凋亡的模型并明确了凋亡启动不可逆时相。如图 14-1 所示的生长曲线，在 10^{-6}mol/L ATRA 诱导 HL-60 细胞的第 1～2 天，不能抑制 HL-60 细胞的生长（生长速率分

别降低 3.1% 和 4.8%），但在第3～4 天生长速率很明显降低 17.1% 和 40.0%，特别在第 5 天时，降低了 61.7%。结果表明，ATRA 以时间依赖的方式抑制细胞的生长。进而，在 ATRA 诱导 HL-60 细胞的第 1、2 和 3 天凋亡速率达 8.5%，与对照组未见显著差异（8.3%），在第 4、5 天时明显增至 14.5% 和 27.4%，而在第 5 天凋亡速率是对照组细胞的 3 倍。同时，流式细胞检测也表明，在诱导的第 5 天时，出现明显的凋亡。因此，凋亡执行的高峰在 ATRA 作用的第 5 天，并且，从第 3 天开始，细胞出现生长抑制。

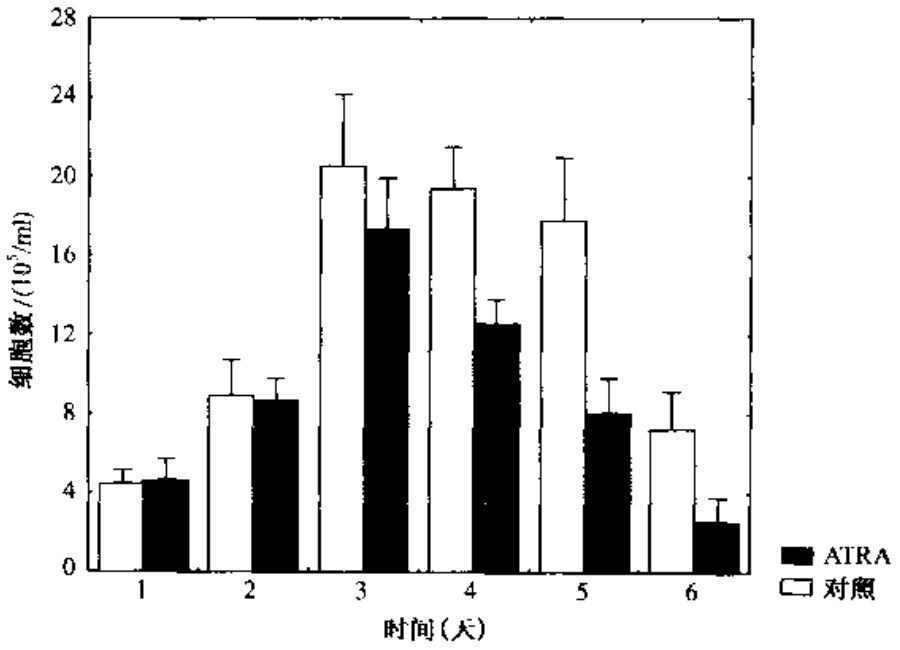


图 14-1　ATRA 诱导 HL-60 细胞的生长曲线

为了发现决定细胞凋亡启动命运的关键蛋白，我们进一步实行 ATRA 撤药实验来明确凋亡启动的时相。首先，当 10^{-6}mol/L ATRA 加入培养基 1～5 天时，分别进行撤药实验（即加入 ATRA 后，每隔一段时间更换新鲜的培养液，继续培养至第六天）。在第六天通过透射电镜和流式细胞仪观察凋亡的情况。流式细胞结果表明，撤药 3 天组的凋亡细胞的百分比（18.1%）明显高于撤药 1、2 天组和对照组（分别为 7.3%，10.0%和 8.4%），并随着时间的延长，凋亡的程度增加。同时，透射电镜观察发现，在 ATRA 撤药的第 3、5 天，凋亡的细胞核出现浓缩，碎裂成几块，细胞形成凋亡小体（图 14-2）。这些结果表明，在 ATRA诱导 HL-60 的第 3 天是凋亡启动的时相。因此，凋亡启动的细胞模型建立起来。

如果凋亡启动的模型已经建立起来，那么，几个已知与凋亡相关的蛋白被用来进一步验证这个模型。Western Blot 结果表明（图 14-3），仅在 ATRA 诱导的

第 3 天，即凋亡启动时相时，STAT3 表达增加直至第 5 天；而磷酸化的 STAT3 与之相反，在第 3 天表达开始下调，此外，拮抗凋亡的 Bcl-2 蛋白自第 3 天开始出现下调，而凋亡的执行者——caspase3 在 ATRA 诱导第 3 天时出现上调。这些结果进一步证实了上述所描述的凋亡启动模型，并且也表明了这些信号转导子在 ATRA 诱导的凋亡启动中发挥一定的作用。

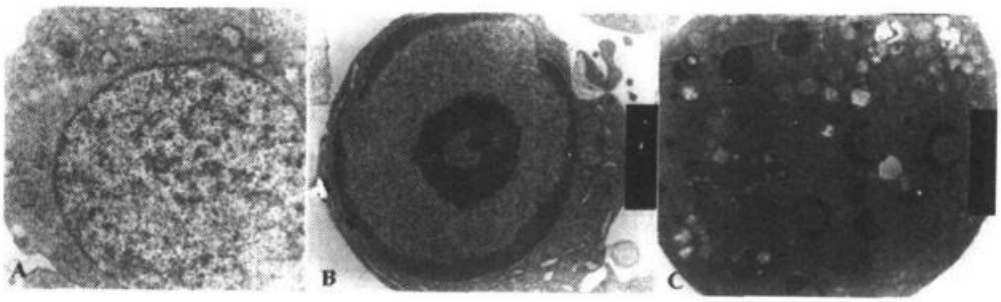


图 14-2　ATRA 诱导 HL-60 细胞撒药实验的透射电镜显示

A. 对照组。B. ATRA 撒药实验诱导第 3 天组。

C. ATRA 撒药实验诱导第 5 天组。

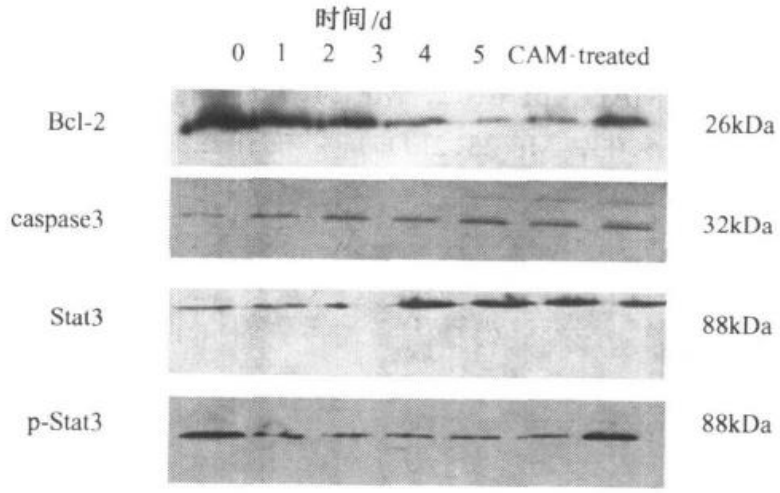


图 14-3　Western Blot 的结果显示与凋亡相关蛋白在 ATRA 诱导 HL-60 细胞凋亡过程中的变化

为了更多地揭示与凋亡启动相关的蛋白以寻找潜在的药物靶点，我们利用比较蛋白质组学的方法，分离比较 ATRA 诱导的凋亡启动 HL-60 细胞和对照组细胞的总蛋白，利用 PDQuest 软件进行计算机图像分析。用 pH＝3～10NL 和 pH＝3～10L 范围的胶条进行等电聚焦，考马斯亮蓝染色可检测到 800 个蛋白斑点（图 14-4）。对照组的配比率为 77.97±1.19%，凋亡启动组的配比率为 80.00±2.29%。对照组的相关系数 correlation coefficient（CC）为 0.7502±0.0190，而凋亡启动组的 CC 为 0.7520±0.0056，两者间的 CC 为 0.7474±0.0290。进而进行

定量的比较（较对照组高 2 倍的数值），一些蛋白分布在偏碱性区域，而另一些斑点分配在中性区域。其中，在凋亡启动组，利用 MALDI-TOF，通过肽质量指纹图谱，查寻 PepIdent 数据库鉴定出 12 个蛋白上调和 7 个蛋白下调。12 个上调蛋白是丙酮酸激酶，M2 同工酶（KPY2-HUMAN，pI7.95，Mr 57.8kDa）；磷酸甘油酸变位酶，脑型（PMGB-HUMAN，pI 6.75，Mr 28.7kDa）；三磷酸甘油醛脱氢酶，肝型（G3P2-HUMAN，pI 8.58，Mr 35.9kDa）；硫氧还蛋白过氧化物酶 2（TDX2-HUMAN，pI 8.27，Mr 22.1kDa）；肌动蛋白，胞质型 1（ACTB-HUMAN -1，pI 5.22，Mr 41.6kDa）；cofilin（COF1-HUMAN，pI 8.22，Mr 18.5kDa）；SM22α homolog（SM2H-HUMAN，pI 8.41，Mr 22.4kDa）；核内不均一核糖核蛋白 A2/B1（ROA2-HUMAN，pI 8.97，Mr 37.4kDa）；鸟嘌呤核苷酸结合蛋白 G（I）/G（S）/G（T）β 亚单位 2（GBB2-HUMAN（pI 5.60，Mr 37.3kDa）；GTP 结合蛋白（pI 7.00，Mr 27.4kDa）；电子转运黄素 β-亚单位（ETFB-HUMAN，pI 8.25，Mr 27.8kDa）和雌激素受体（ESR1-HUMAN，pI 7.00，Mr 53.7kDa）。7 个下调蛋白是 poly（RC）-结合蛋白 1（HnRNP E1）（PCB1-HUMAN，pI 6.66，Mr 37.5kDa）；鸟氨酸转移酶（OAT-HUMAN-1，pI 5.72，Mr 47.8kDa）；高迁移蛋白 HMG1（HMG1-HUMAN，pI 5.62，Mr 24.7kDa）；蛋白酶体亚单位 α3 型（PSA3-HUMAN，pI 5.19，Mr 28.3kDa）；抗氧化蛋白 2（AOP2-HUMAN，pI 6.02，Mr 24.9kDa）；Rho GDP-解聚抑制因子 1（GDIR-HUMAN，pI 5.03，Mr 23.2kDa）和核苷二磷酸激酶 B（NDKB-HUMAN，pI 6.97，Mr 17.4kDa）。

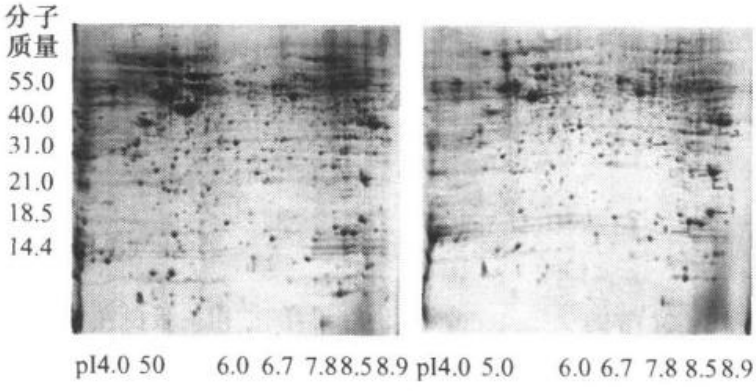


图 14-4 ATRA 诱导的 HL-60 细胞凋亡启动的双向电泳图谱
（考马斯亮蓝染色）

目前，凋亡启动就像细胞周期的“检查点”一样，可能参与决定凋亡的不可逆命运。因而，不仅从整体上深入理解凋亡的信号传导是必需的，而且从寻找治疗恶性肿瘤新策略的药靶来讲也是重要的。首先，在这项工作中，我们研究表明，在 ATRA 体外诱导 HL-60 细胞 3d 时，细胞凋亡启动，此时，Bcl-2 和磷酸

化 STAT3 表达下调，而 STAT3 和 caspase3 同时出现上调，并且持续到凋亡执行发生。Bcl-2 家族蛋白已经证实可通过相互作用为细胞设定阈值，同源或异源二聚体间相互结合导致的立体化学变化将启动细胞不可逆死亡[33]。而 Minami M 等报道在髓性白血病细胞中，野生型 STAT3 的过表达加速了 IL-6 或 LIF 所诱导的凋亡[34]。而 STAT3 的反义核酸可有效抑制其激活，并且可增强肿瘤细胞凋亡[35]。另外，在大鼠基质-血管通过 RA（10μmol/L）调节凋亡可导致减少脂细胞[36]。总之，这些特征性分子的改变提示 ATRA 诱导的 HL-60 细胞凋亡启动模型具有一定的客观性和代表性，并且能够利用此模型研究参与凋亡启动的相关分子。

我们利用蛋白质组技术分析 ATRA 体外诱导凋亡启动模型，初步显示一些蛋白表达的状态，并且揭示了与凋亡启动相关的 19 个蛋白（光密度值至少具有 2 倍变化），其中，12 个蛋白在凋亡启动时表达上调，7 个蛋白表达下调。根据它们已知的功能，我们将其分为 5 类：与 G-蛋白和核受体信号转导相关的蛋白、与蛋白质代谢、脂类代谢和糖代谢相关的蛋白、与氧化还原状态相关的蛋白以及与细胞核/细胞骨架相关的蛋白。在 G-蛋白和核受体信号转导相关蛋白这类蛋白质中，GDIR 和 ESR 被鉴定出来。其中，GTP-结合蛋白和 Gβ 表达上调，但 GDIR 表达下调。文献报道，GTP-结合蛋白被认为是细胞死亡的介导子[37]，并且 RA 可以显著增强 GTP-结合蛋白/谷氨酰胺转移酶结合 GTP 的能力[38]。与上述相反，GDIR 的断裂能够激活 Jun N 端激酶，提示在某些系统中它作为凋亡的上游调节因子存在[39]。所有这些结果表明，与 G-蛋白信号转导相关的蛋白与 RA 诱导 HL-60 细胞凋亡启动相关。此外，有些文献表明 ESR 诱导的凋亡与 p38 MAPK 的激活途径相关[40]。ERβ 曾介导神经细胞的凋亡，而 ERα 则具有保护作用[41]。在我们的实验中，ERα 在 HL-60 细胞凋亡启动的过程中表达上调，提示 ERα 可能在不同的细胞类型中发挥不同的作用。

在蛋白质代谢中，PSA3 和 OAT 在 RA 诱导的凋亡启动中表达下调。文献报道，泛素-蛋白酶体依赖的蛋白降解系统在许多细胞模型中已经显示参与调节凋亡。通过加入抑制蛋白酶体功能的 lactacystin，慢性 B 型淋巴细胞白血病对于凋亡敏感性增高[42]。此外，目前未见任何关于与氨基酸合成相关的 OAT 与凋亡的文献报道。

KPY2 和 PMGB 在糖代谢途径的不同阶段发挥重要的作用。两者在凋亡启动的过程中表达皆上调。目前，对糖代谢与凋亡之间的关系知之甚少。然而，已知 KPY2 和 PMGB 是肝细胞和睾丸组织的分化标志物[43,44]。Martin 等也表明 A-TRA 诱导的 HL-60 细胞首先经过分化，随后在体外发生凋亡[45]。因此，在我们的实验模型中，KPY2 和 PMGB 的上调提示 ATRA 触发的 HL-60 细胞的分化可能是凋亡发生的前提。

ETFB 和 HnRNP-E1 与脂类代谢相关。到目前为止，没有文献报道这些蛋白

与凋亡相关。在 RA 诱导的 HL-60 细胞凋亡启动过程中，ETFB 表达上调，而 HnRNP-E1 表达下调。Antonacci 等研究表明野生型 ETFB 的 cDNA 能够弥补遗传的缺陷，恢复 β-氧化至正常水平[46]。这些数据提示与脂类代谢相关的 β-氧化可能参与凋亡启动，提供能量而防止凋亡。此外，文献报道，结合 HnRNP-E1 的 mRNA 可以调节由 12/15-lipoxygenases（LOX）I 催化的脂类过氧化[47]。故推测 HnRNP-E1 的下调可能增加脂质过氧化，并且在凋亡启动中保持能量。

TDxII 和 AOP2 与细胞的氧化还原相关。在凋亡启动中，TDxII 表达上调，而 AOP2 表达下调。近来文献证实，在 Molt-4 白血病细胞中稳定过表达的 TDxII 能够保护因血清撤除所致的凋亡，表明它具有通过限制破坏性的氧自由基来防止氧化应激的功能[48]，因此，TDxII 在凋亡的启动过程中，被认为是细胞防御的调节因子。AOP2 是巯基特异性抗氧化家族的成员。尽管没有关于它与凋亡的相关报导，但是另一个抗氧化物质 pyrrolidine dithiocarbamate（PDTC）可明显防止由 TNFα 所诱导的凋亡[49]。这些数据提示 AOP2 很可能通过某些途径来防止凋亡的发生，其下调有可能在凋亡启动中发挥某些积极的作用。

在我们的研究中，与细胞核相关的三个蛋白被鉴定出来。NDKB 和 HMG1 在凋亡启动中表达下调。Fung 等报道 NDKB 在恶性 NPE 细胞中过表达[50]，能刺激 c-Myc 的转录[51]。因此，在凋亡启动中其下调很可能通过 c-Myc 的调节来发挥作用。而 HMG1 是一个非组蛋白，与染色体相关。Romine 等报道 HMG1 可能增强核受体与其反应元件的结合，并且与共激活因子相互作用来调节激素反应基因[52]。尽管没有关于 HMG1 与凋亡相关的报道，但已知 HMG2 接受信号表达增加，并可直接诱导细胞核形态发生凋亡变化和 DNA 断裂[53]。这些数据提示 HMG1 可能在凋亡启动中发挥一定的作用。

actin 在凋亡启动中表达上调。曾报道作为细胞骨架蛋白的 actin 的破坏和重组与凋亡小体的形成相关[54]。此外，其多聚化能改变在线粒体水平早期发挥作用的 gelsolin 的抑制活性，最后增强凋亡[55]。因此，actin 作为凋亡启动的阳性调节因子存在是可能的。同样，解聚蛋白 cofilin 也在凋亡启动中表达增加。文献报道防止 cofilin 的去磷酸化和核易位可导致恶性 T 淋巴细胞的凋亡[56]。因此，这些数据提示 actin 和 cofilin 的增加可能与凋亡启动相关。

SM2H 和 HnRNPA2/B1 也在凋亡启动中出现。尽管没有直接证据表明它们与凋亡相关，但各种数据确实揭示 SM2H 和 HnRNPA2/B1 所属的 calpnin 和 HnRNP 家族其他成员在凋亡中可发挥一定的作用[57]。因此，推测它们也可能参与凋亡启动。

因此，在 ATRA 诱导 HL-60 细胞凋亡启动的蛋白质组研究中，某些蛋白可能作为触发凋亡的阳性调节因子，而其他则作为反应性的抑制凋亡蛋白存在。这表明在细胞凋亡启动的过程中可能存在一个微妙的平衡，一旦平衡被打破，则细胞必将走向凋亡（图 14-5）[58]。

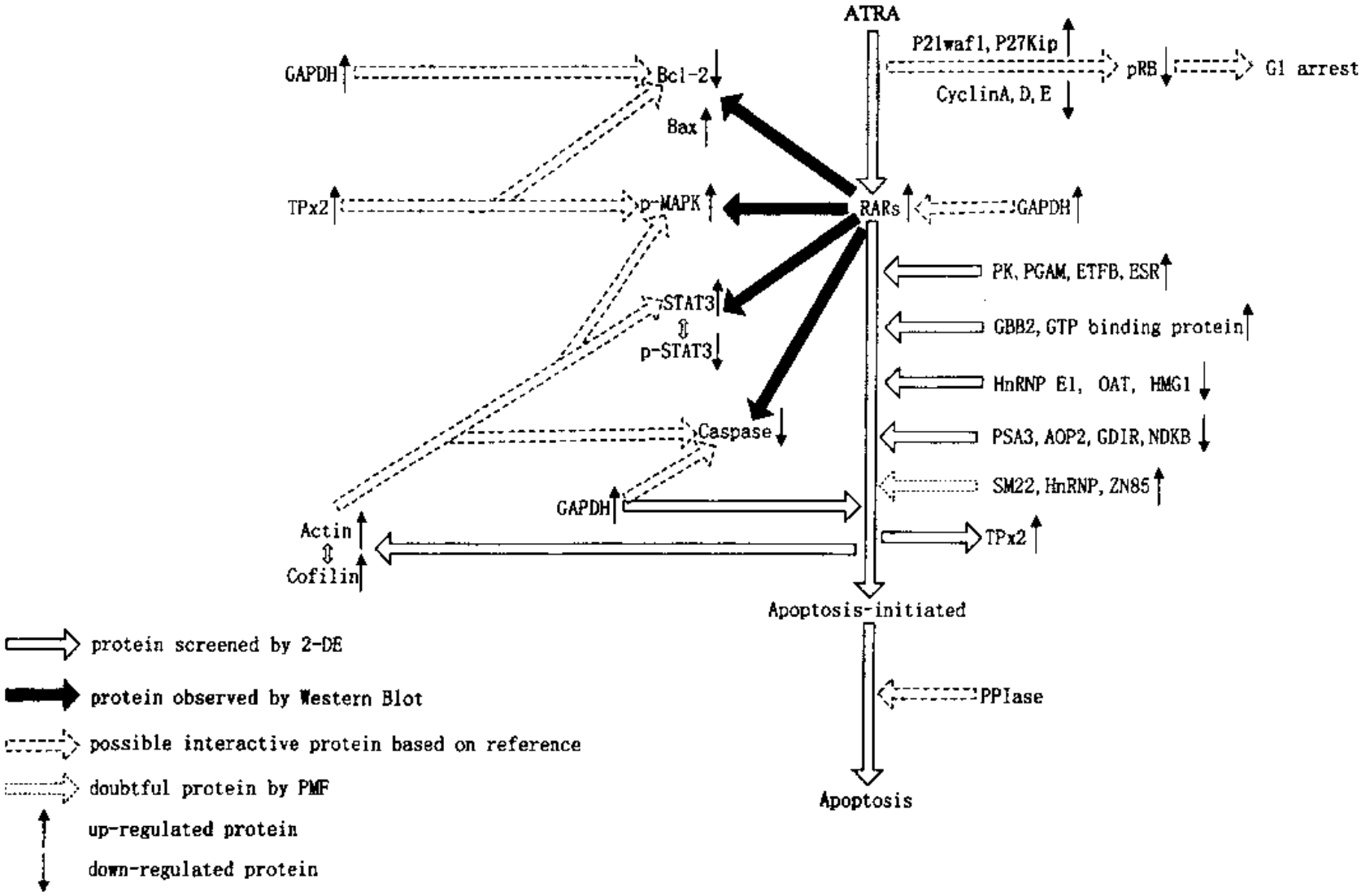

ATRA
P21waf1, P27Kip
pRB
G1 arrest
CyclinA, D, E
GAPDH
Bcl-2
Bax
TPx2
p-MAPK
RARs
GAPDH
PK, PGAM, ETFB, ESR
STAT3
p-STAT3
GBB2, GTP binding protein
HnRNP E1, OAT, HMG1
Caspase
PSA3, AOP2, GDIR, NDKB
SM22, HnRNP, ZN85
GAPDH
Actin
Cofilin
TPx2
Apoptosis-initiated
PPIase
Apoptosis
protein screened by 2-DE
protein observed by Western Blot
possible interactive protein based on reference
doubtful protein by PMF
up-regulated protein
down-regulated protein

图 14-5

综上所述，为了获得诱导细胞发生凋亡的全面认识，深入地了解其复杂机制，采用蛋白质组学的方法来研究凋亡问题是必需的。

参考文献

1 Liang P, Pardee A B. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science, 1992, 257: 967～971

2 Wasinger V C, Cordwell S J, Cerpa-Poljak A, Yan J X, Gooley A A, Wilkins M R, Duncan M W, Harris R, Williams K L, Humphery-Smith I. Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis, 1995, 16 (7): 1090-1094

3 贺福初. 蛋白质组（Proteome）研究——后基因组时代的生力军. 科学通报, 1999, 44 (2): 113～122

4 Prasad S C, Soldatenkov V A, Kuettel M R, Thraves P J, Zou X, Dritschilo A. Protein changes associated with ionizing radiation-induced apoptosis in human prostate epithelial tumor cells. Electrophoresis, 1999, 20: 1065～1074

5 胡志远, 贺福初. 蛋白质组研究进展. 生物化学与生物物理进展, 1999, 26 (3): 202～205

6 Jensen L M, Zhang Y, Shooter E M. Steady-state polypeptide modulations associated with nerve growth factor (NGF) -induced terminal differentiation and NGF deprivation-induced apoptosis in human neuroblastoma cells. J Biol Chem, 1992, 267 (27): 19325～19333

7 Villa P, Miehe M, Sensenbrenner M, Pettmann B. Synthesis of specific proteins in trophic factor-deprived neurons undergoing apoptosis. J Neurochem, 1994, 62 (4): 1468～1475

8 Amess B, Tolkovsky I M. Programmed cell death in sympathetic neurons: a study by two-dimensional polyacrylamide gel electrophoresis using computer image analysis. Electrophoresis, 1995, 16: 1255～1267

9 Robaye B, Doskeland A P, Suarez-Huerta N, Doskeland S O, Dumont J E. Apoptotic cell death analyzed at the molecular level by two-dimensional gel electrophoresis. Electrophoresis, 1994, 15: 503～510

10 Prasad S C, Soldatenkov V, Notario V, Smulson M, Dritschilo A. Detection of heterogeneneity of apoptotic fragments of poly (ADP-ribose) polymerase in MDA-MB-468 breast cancer cells: two-dimensional gel analysis. Electrophoresis, 1999, 20: 618～625

11 Griffiths G J, Dubrez L, Morgan C P, Jones N A, Whitehouse J, Corfe B M, Dive C, Hickman J A. Cell damage-induced conformational changes of the pro-apoptotic protein Bak in vivo precede the onset of apoptosis. J Cell Bio, 1999, 144 (5): 903～914

12 Muller E C, Schumann M, Rickers A, Bommert K, Wittmann-Liebold B, Otto A. Study of Burkitt Lymphoma cell line proteins by high resolution two-dimensional gel electrophoresis and nanoelectrophospray mass spectrometry. Electrophoresis, 1999, 29: 320～330

13 Rickers A, Brockstedt E, Mapara M Y, Otto A, Dorken B, Bommert K. Inhibition of CPP32 blocks surface IgM-mediated apoptosis and D4-GDI cleavage in human BL60 Burkitt lymphoma cells. Eur J Immunol, 1998, 28 (3): 1122

14 Robaye B, Doskeland A P, Suarez Huerta N, Doskeland S O, Dumont J E. Apoptotic cell death analyzed at the molecular level by two-dimensional gel electrophoresis. Electrophoresis, 1994, 15 (3-4): 503～510

15 Yoo B C, Vikolinsky R, Engidawork E, Cairns N, Fountoulakis M, Lubec G. Differential expression of molecular chaperones in brain of patients with Down syndrome. Electrophoresis, 2001, 22 (6): 1233～1241

16 Araki N, Morimasa T, Sakai T, Tokuoh H, Yunoue S, Kamo M, Miyazaki K, Abe K, Saya H, Tsugita A.

Comparative analysis of brain proteins from p53-deficient mice by two-dimensional electrophoresis. Electrophoresis, 2000, 21 (9): 1880～1889

17 Patterson S D, Spahr C S, Daugas E, Susin S A, Irinopoulou T, Koehler C, Kroemer G. Mass spectrometric identification of proteins released from mitochondria undergoing permeability transition. Cell Death Differ, 2000, 7 (2): 137～144

18 Gerner C, Frohwein U, Gotzmann J, Bayer E, Gelbmann D, Bursch W, Schulte-Hermann R. The Fas-induced apoptosis analyzed by high throughput proteome analysis. J Biol Chem, 2000, 275 (50): 39018～39026

19 Thiede B, Dimmler C, Siejak F, Rudel T. Predominant identification of RNA-binding proteins in Fas-induced apoptosis by proteome analysis. J Biol Chem, 2001, 276 (28): 26044～26050

20 Suzuki T, Terasaki M, Takemoto-Hori C, Hanada T, Ueda T, Wada A, Watanabe K. Proteomic analysis of the mammalian mitochondrial ribosome. Identification of protein components in the 28 S small subunit. J Biol Chem, 2001, 276 (35): 33181～33195

21 Thery C, Boussac M, Veron P, Ricciardi-Castagnoli P, Raposo G, Garin J, Amigorena S. Proteomic analysis of dendritic cell-derived exosomes: a secreted subcellular compartment distinct from apoptotic vesicles. J Immunol, 2001, 166 (12): 7309～7318

22 Macdonald N, Barrow K, Tonge R, Davison M, Roberts R A, Chevalier S. PPARalpha-dependent alteration of GRP94 expression in mouse hepatocytes. Biochem Biophys Res Commun, 2000, 277 (3): 699～704

23 Olins A L, Buendia B, Herrmann H, Lichter P, Olins D E. Retinoic acid induction of nuclear envelope-limited chromatin sheets in HL-60. Exp Cell Res, 1998, 245: 91～104

24 Martin S J, Bradley J G, Cotter T G. HL-60 cells induced to differentiation towards neutrophils subsequently die via apoptosis, Clin Exp Immunol, 1990, 70: 448～453

25 Mangelsdorf D J, Evans R M, The RXR heterodimers and orphan receptors. Cell, 83 (6): 841～850

26 Park J R, Robertson K, Hickstein D D, Tsai S, Hockenbery D M, Collins S J. Dysregulated Bcl-2 expression inhibits apoptosis but not differentiation of retinoic acid-induced HL-60 granulocytes, Blood, 1994, 84 (2): 440～445

27 Watson R W, Rotstein O D, Parodo J, Bitar R, Hackam D, Marshall J C. Granulocytic differentiation of HL-60 cells results in spontaneous apoptosis mediated by increased caspase expression, FEBS Leters, 1997, 412: 603～609

28 Vincenti M P, Schroen D J, Coon C I, Brinckerhoff C E. v-src activation of the collagenase-1 (matrix metalloproteinse-1) promoter through PEA3 and STAT: requirement of extracellular signal regulated kinases and inhibition by retinoic acid receptors, Mol Carcinog, 1998, 21 (3): 194～204

29 Joseph B, Lefebvre O, Mereau-Richard C, Danze P M, Belin-Plancot M T, Formstecher P. Evidence for the involvement of both retinoic acid receptor-and retinoic X receptor-dependent signaling pathways in the induction of tissue tranglutaminase and apoptosis in the human myeloma cell line RPMI8226. Blood, 1998, 9 (7): 2423～2432

30 Fujimura S, Suzumiya J, Anzai K, Ohkubo K, Hata T, Yamada Y, Kamihira S, Kikuchi M, Ono J. Retinoic acids induce growth inhibition and apoptosis in adult T-cell leukemia (ATL) cell lines, Leuk Res, 1998, 22 (7): 611～618

31 Okamura T, Masuda M, Arai Y, Ishida C, Shudou K, Mizoguchi H. All-trans retinoic acid modulates Fas antigen expression and affects cell proliferation and apoptosis in combination with anti-Fas monoclonal antibody in the human myeloma cell line, U266B1. Exp Hematol, 1998, 26 (6): 501～506

32 Mantymaa P, Guttorm T, Siitonen T, Saily M, Savolainen E R, Levonen A L, Kinnula V, Koistinen P.

Cellular redox state and its relationship to the inhibition of clonal cell growth and the induction of apoptosis during all-trans retinoic acid exposure in acute myeloblastic leukemia cells. Haematologica, 2000, 85 (3): 238－245

33 Griffiths G J, Dubrez L, Morgan C P, Jones N A, Whitehouse J, Corfe B M, Dive C, Hickman J A. Cell damage-induced conformational changes of the pro-apoptotic protein Bak in vivo precede the onset of apoptosis. J Cell Biol, 1999, 144 (5): 903－914

34 Minami M, Inoue M, Wei S, Takeda K, Matsumoto M, Kishimoto T, Akira S. STAT3 activation is a critical step in gp130-mediated terminal differentiation and growth arrest of a myeloid cell line. Proc Natl Acad Sci USA, 1996, 93: 3968－3976

35 Grandis J R, Drenning S D, Zeng Q, Watkins S C, Melhem M F, Endo S, Johnson D E, Huang L, He Y, Kim J D. Constitutive activation of Stat3 signaling abrogates apoptosis in squamous cell carcinogenesis in vivo. Proc Natl Acad Sci USA, 2000, 97 (8): 4227－4232

36 Kim H S, Hausman D B, Compton M M, Dean R G, Martin R J, Hausman G J, Hartzell D L, Baile C A. Induction of apoptosis by all trans-retinoic acid and C2-ceramide treatment in rat stromal-vascular cultures. Biochem Biophys Res Commun, 2000, 270 (1): 76－80

37 Turner P R, Mefford S, Christakos S, Nissenson R A, Apoptosis mediated by activation of the G protein-coupled receptor for *parathyroid hormone* (PTH) /*PTH-related protein* (PTHrP). Mol Endocrinol, 2000, 14 (2): 241－254

38 Singh U S, Cerione R A. Biochemical effects of retinoic acid on GTP-binding Protein/Transglutaminases in HeLa cells. Stimulation of GTP-binding and transglutaminase activity, membrane association, and phosphatidylinositol lipid turnover. J Biol Chem, 1996, 271 (44): 27292－27298

39 Krieser R J, Eastman A. Cleavage and nuclear translocation of the caspase 3 substrate *Rho* GDP-dissociation inhibitor, D4-GDI, during apoptosis. Cell Death Differ, 1999, 6 (5): 412－419

40 Zhang C C, Shapiro D J, Activation of the p38 mitogen-activated protein kinase pathway by estrogen or by 4-hydroxytamoxifen is coupled to estrogen receptor-induced apoptosis. J Biol Chem, 2000, 275 (1): 479－486

41 Okamoto K, Andreola F, Chiantore M V, Dedrick R L, De Luca L M, Differences in uptake and metabolism of retinoic acid between estrogen receptor-positive and -negative human breast cancer cells. Cancer Chemother Pharmacol, 2000, 46 (2): 128－134

42 Masdehors P, Merle-Beral H, Magdelenat H, Delic J. *Ubiquitin-proteasome system and increased sensitivity* of B-CLL lymphocytes to apoptotic death activation. Leuk Lymphoma, 2000, 38 (5-6): 499－504

43 Guguen-Guillouzo C, Szajneri M F, Glaise D, Gregori C, Schapira F. Isozyme differentiation of aldolase and pyruvate kinase in fetal, regenerating, preneoplastic, and malignant rat hepatocytes during culture. In Vitro, 1981, 1795: 369－377

44 Omenn G S, Cheung S C, Phosphoglycerate mutase isozyme marker for tissue differentiation in man. Am J Hum Genet, 1974, 26 (3): 393－399

45 Martin S J, Bradley J G, Cotter T G. HL-60 cells induced to differentiate towards neutrophils subsequently die *via apoptosis*. *Clin Exp Immunol*, 1974, 70: 448－453

46 Antonacci R, Colombo I, Archidiacono N, Volta M, DiDonato S, Finocchiaro G, Rocchi M. Molecular cloning, cDNA sequence, and chromosomal localization of the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene. Genomics, 1994, 19 (1): 177－179

47 Ostareck D H, Ostareck-Lederer A, Wilm M, Thiele B J, Mann M, Hentze M W. mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation *from the 3' end*. *Cell*,

1997, 89 (4): 597～606

48 Ferret P J, Soum E, Negre O, Wollman E E, Fradelizi D. Protective effect of thioredoxin upon NO-mediated cell injury in THP1 monocytic human cells. Biochem J, 2000, 346: 759～765

49 Lee D H, Macintyre J P, Taylor G R, Wang E, Plante R K, Tam S S, Pope B L, Lau C Y. Inhibition of activation of nuclear factor kappaB is responsible for inhibition of inducible nitric oxide synthase expression by higenamine, an active component of aconite root. J Pharmacol Exp Ther, 1999, 289 (3): 1465～1471

50 Fung L F, Lo A K, Yuen P W, Liu Y, Wang X H, Tsao S W. Differential gene expression in nasopharyngeal carcinoma cells. Life Sci, 2000, 7 (8): 923～936

51 Agou F, Raveh S, Mesnildrey S, Veron M. Single strand DNA specificity analysis of human nucleoside diphosphate kinase B. J Biol Chem, 1999, 274 (28): 19630～19638

52 Romine L E, Wood J R, Lamia L A, Prendergast P, Edwards D P, Nardulli A M. The high mobility group protein 1 enhances binding of the estrogen receptor DNA binding domain to the estrogen response element. Mol Endocrinol, 1998, 12 (5): 664～674

53 Kawakami T, Nagata T, Muraguchi A, Nishimura T. Alteration of protein composition in mouse thymocytes by signals through T-cell receptor. Electrophoresis, 2000, 21 (9): 1846～1852

54 Suarez-Huerta N, Mosselmans R, Dumont J E, Robaye B. Actin depolymerization and polymerization are required during apoptosis in endothelial cells. J Cell Physiol, 2000, 184 (2): 239～245

55 Koya R C, Fujita H, Shimizu S, Ohtsu M, Takimoto M, Tsujimoto Y, Kuzumaki N. Gelsolin inhibits apoptosis by blocking mitochondrial membrane potential loss and cytochrome c release. J Biol Chem, 2000, 275 (20): 15343～15349

56 Samstag Y, Dreizler E M, Ambach A, Sczakiel G, Meuer S C. Inhibition of constitutive serine phosphatase activity in T lymphoma cells results in phosphorylation of pp19/cofilin and induces apoptosis. J Immunol 1996, 156: 4167～4173

57 Muller E C, Schumann M, Rickers A, Bommert K, Wittmann-Liebold B, Otto A, Study of Burkitt lymphoma cell line proteins by high resolution two-dimensional gel electrophoresis and nanoelectrospray mass spectrometry. Electrophoresis, 1999, 29: 320～330

58 Wan J, Wang J, Cheng H, Yu Y, Xing G, Qiu Z, Qian X, He F. Proteomic analysis of apoptosis initiation induced by all-trans retinoic acid in human acute promyelocytic leukemia cells. Electrophoresis, 2001, 22 (14): 3026～3037

（万晶宏）

第十五章　蛋白质组与新药开发

目前世界上存在的已知疾病大约有4500种，其中只有不到二分之一的疾病有相应的治疗手段，可以有药物对应的疾病甚至不到四分之一。一个新药从开发到上市，历经十几年的时间，研究开发经费可达十几亿美元。可是，一旦上市以后，它所带来的利润也是相当可观的。可见新药的研究开发具有广阔的前景，是一个周期长、高风险、高投入、高回报的产业。当代创新药物研究竞争十分剧烈，其中竞争的焦点就在于筛选新药，问题的核心是低成本、高效率地筛选出新药，以达到缩短新药发现过程的目标。在医药工业，发现一个新药后，临床前药理、毒理学研究平均需要3.5年，临床试验平均需要6年，审批上市平均需要2.5年，三者相加平均12年（如图15-1）[1,2]。企图从这几个发展阶段上缩短周期，加快新药上市的速度是很困难的。这几个阶段周期幅度，不但受新药研究规律所制约，也受药品审批管理法规要求的限制。

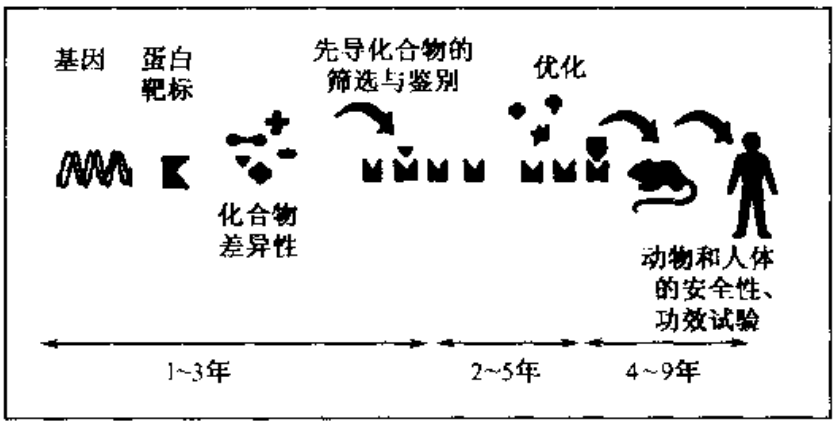


图15 1　新药发现的过程

针对上述难题，新药研制机构，特别是各国的制药企业公司纷纷加大人力、财力的投入力度，将热情和注意力倾注在新药发现过程上，大量使用现代新技术，寻找切实可行的解决方案，以期抢占制药行业的先机。这就在很大程度上促进了相关科技的高速发展。基因组学（genomics）、组合化学（combinatorial chemistry）和高通量筛选（high-throughput screening，HTS）等三种技术的联合应用使得新药研制的成本得到大幅度下降，加速了具有特定靶标的新化合物的发

现。人类基因组测序工作的完成，也将为药物的寻找提供更多的可供筛选的靶标。

开始于1990年的HGP（human genome project），是一个大型的国际研究合作项目，人类基因组的全部序列已于2001年2月份公布。单核苷酸多态性（single nucleotide polymorphisms，SNP）是指在基因组DNA某个位置存在单个碱基的差异，人类基因组DNA序列90%多态性是由它造成的，个体对环境和治疗的应答呈多样性，其根源就在于SNP。人类基因组中大约有300万SNP直接与疾病有关或影响药物处治的药代动力学。其研究已成为当前功能基因组研究中的一个新的热点领域。在药物处治过程中，人体的每种细胞型大约有100种基因显示出明显不同的表达结果[3]。

在欧美一些国家，将HGP的研究和新药开发结合在一起，创立了一门新的学科，叫药物基因组学（pharmacogenomics）。它以基因组学为基础，以研究开发新药为目的，在基因水平上对疾病敏感度、药物反应和副作用进行分析和研究，是药理学当中的一个新分支。其大致含义包括：药物效应的基因型预测和基因组学在医药工业上的应用，在分子水平上证明和阐述药物疗效，药物作用的靶标、作用模式和毒副作用。

但是，核酸作为药物作用的靶标实际上存在着致命的缺陷：首先由于核酸结构的同源性，联系到许多人类正常的功能，作用于DNA的药物大多选择性差，而且毒性较大，特别是形成共价结合的药物都有严重的细胞毒作用；其次大多数疾病都是表征在蛋白质水平上，而不是在基因水平上，细胞和组织中mRNA的丰度与蛋白质的丰度相关性不显著（仅为0.5）[4]；第三，将核酸结合的药物释放到相应的组织是一大难题；第四，很多疾病如癌症和心血管疾病往往是多基因共同作用的结果，因此很难找到关键基因作为作用的靶标。以上几点都是制约新药开发进程的瓶颈。

因此，为了阐明生命活动的本质，真正与功能研究结合，基因组的研究必然要回归到蛋白质组研究。蛋白质的结构、分布和功能，对于预测某些疾病的进程、药物作用及其过程，对于阐明不同个体之间遗传学上的差异是非常重要的。

蛋白质组是指基因组表达的所有相应的蛋白质，是细胞或机体全部蛋白质的存在及其活动方式。它是从整体的蛋白质水平上，从生命本质的层次上研究和发现生命活动的规律和重要的生理、病理现象的本质。与mRNA表达分析相似的是，它揭示了基因活性的动态表达，而不是基因组的静态行为。由于疾病的发生与发展、药物的作用大多是在蛋白质水平上进行的，因此蛋白质组学研究克服了蛋白质表达和基因之间的非线性关系。将蛋白质组应用于新药研究，可以通过对疾病与正常细胞中的蛋白质组进行比较，发现可以成为药物筛选的作用靶标的、与一些疾病相关的蛋白质。而新药及其靶标的发现和药物作用的模式研究是药物蛋白质组学的重要研究内容。

为了尽快发掘到具有可成为疾病标志物或药物作用靶标的蛋白质，蛋白质组学技术向高通量（大规模）和自动化方向发展，而高灵敏度、高准确度和高重现性成为评价技术发展的标准。

蛋白质组学可以提供一种发现和鉴定在疾病作用下表达异常的蛋白质的方法。这种蛋白质可以作为药物筛选的作用靶点。还可以通过对疾病发生的不同阶段蛋白质的变化进行分析，发现一些疾病不同时期的蛋白标志物，这样不仅对药物发现具有指导意义，还可形成未来诊断学、治疗学的基础理论。癌症被称为当今人类的一大天敌，因而抗肿瘤药的开发也就成了科学家们研究的重点。现今大多数抗癌药都伴有严重的毒副作用，特别是对晚期癌症病人进行单一化疗或联合放疗、过热疗法时，经常伴随着癌细胞对细胞抑制剂耐药性的发生。如果能发现在耐药细胞系中表达异常的蛋白质或者与细胞毒性密切相关的蛋白质，就可以此蛋白为靶点设计用于联合用药的新化学单位（new chemical entities)。也可以此信息为参考，设计避免产生耐药性或毒副作用的药物。另外造成癌症病人死亡率极高的主要原因是肿瘤的转移。现在国内一些实验室已开始利用蛋白质组学技术，通过对高低转移细胞株蛋白质的比较，来寻找与肿瘤转移相关的蛋白质。同样也可以高转移株中特异表达的蛋白质为靶点，开发抑制肿瘤转移的新药。

在抗生素方面，由于传染病仍是死亡的主要原因，因而抗感染药是近年来各国新药开发的热点之一。但面对抗生素耐药问题以及不断出现新的微生物的感染性疾病，老的方法显得束手无策。其根本原因就在于对药物的作用机制缺乏透彻的认识。蛋白质组学技术可以让人们清楚地了解细菌内哪些蛋白质会在抗生素的作用下发生改变，以及发生何种变化。根据这些变化，并以蛋白质作为新药设计的靶点，筛选出新一类的抗生素。同时还可选取一些耐药菌株，考察其耐药机制。

第一节　药物分子靶标的高通量筛选体系

在药物研发中，疾病靶标有两个含义，即靶基因和靶蛋白。通常，我们首先需要了解影响某一疾病的基因（整个代谢途径上的基因），其中一个或一系列基因就称为靶基因。靶基因确定之后，就有可能针对与此疾病相关的某一个基因产物——蛋白质设计药物。这种蛋白质就确定为靶蛋白。

目前，已有许多新技术新方法可加快这一阶段的研究，如生物信息学（bioinformatics)、体外或体内表达技术（*in vivo* or *in vitro* expression)、超高通量（ultra high throughput）仪器、微阵列技术（microarrays)、反义技术（anti-sense）以及蛋白质组学（proteomics）技术等等，这些技术是药物早期开发成功的关键。

由于许多疾病与信号传导途径异常有关，因而信号分子可以作为治疗药物设

计的靶点。在信号传递过程中涉及成百上千个蛋白质，蛋白质-蛋白质相互作用发生在细胞内信号传递的所有阶段。而且，这种复杂的蛋白质作用的串联效应可以完全不受基因调节而自发地产生[5]。通过与正常细胞作比较，掌握与疾病细胞中某个信号途径活性增加或丧失有关的蛋白质的改变，将为药物设计提供更为合理的靶点。

基因工程、基因大规模序列分析、DNA 和多肽自动合成、色谱-光谱联用、多维色谱及光谱、组合化学、蛋白质相互作用、高通量及超高通量药物筛选等技术已广泛应用于生命科学的许多领域，特别是新药研究和开发。

生物芯片（biochip）技术是近几年才发展起来的一种高通量、微型化和自动化检测技术，又称微阵列（microarray）分析技术。它应用于生命科学的许多领域，如疾病相关基因的发现、疾病分子诊断和预测及基因功能研究等（见图 15-2)[6]。在药物领域，对于药物靶标的发现、多靶位同步超高通量药物筛选、药物作用的分子机理、中医药基础理论现代化、药物活性及毒性评价等方面都有其他方法无可比拟的优越性。

生物芯片技术至少由 5 部分组成（见图 15-3)[7]：① 具有特殊表面的芯片本身。② 在芯片上点布或原位合成核酸（或蛋白质）探针得到微阵列的设备。布阵的质量直接影响微阵列分析的结果，目前喷泡（bubble jet）技术[8]和寡核苷酸的无屏蔽原位（maskless in situ）合成技术[9]的建立使微阵列分析技术得到很大提高。③ 用于与靶 DNA 杂交的流体系统。④ 读取芯片的扫描仪。⑤ 定量分析和解释结果的完善的软件程序。另外还需要从生物材料中提取核酸以备分析的工具。

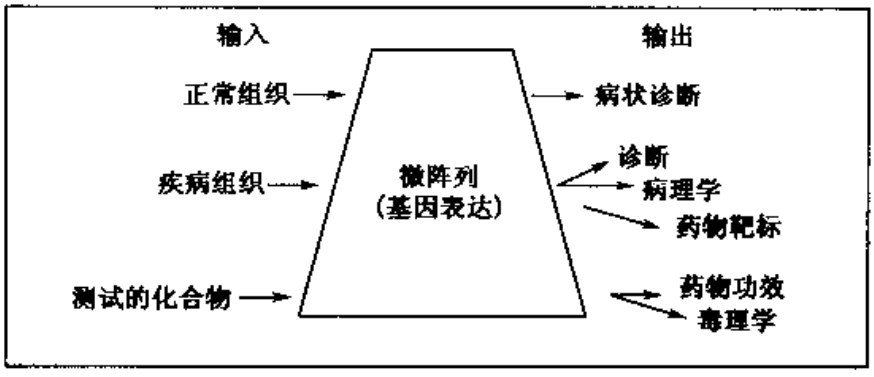


图 15-2　生物芯片为功能基因组学提供综合平台

基因芯片（genechip）及蛋白质芯片（proteinchip）均属于生物芯片，都是近年来发展起来的新的生物技术。现在它们越来越多地应用到生物学、医学、遗传学、药理学和毒理学等多个方面，并开始崭露头角。

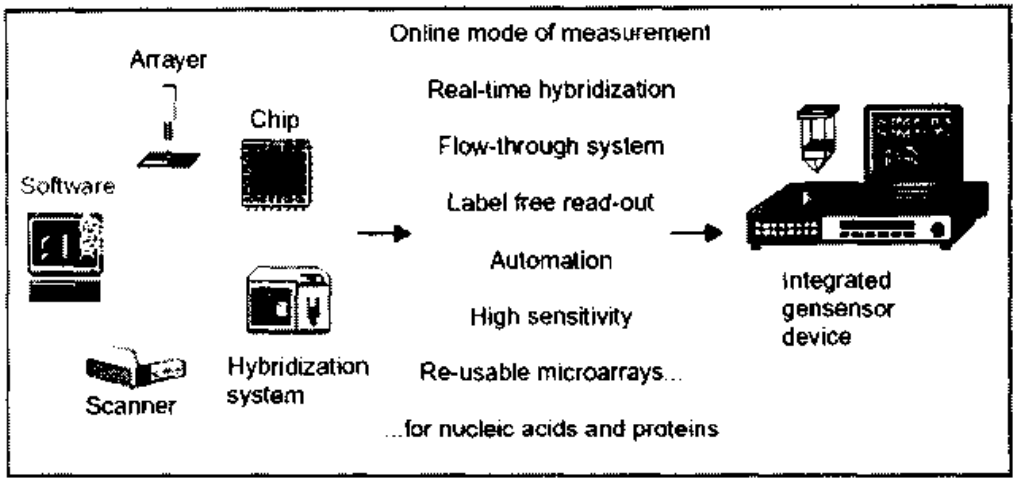


图 15-3　生物芯片及其用于工业化的主要技术[7]

一、基因芯片及其应用

基因芯片又称 DNA 或 cDNA 微阵列（DNA or cDNA microarray），是指在面积很小的载体上，点布数以万计不同的寡核苷酸或 cDNA，将芯片与标记有荧光染料的待测 DNA 或 mRNA 杂交，与靶序列配合好的探针会产生强烈的杂交信号，如果有碱基的错配，信号就会减弱。籍此能判断靶 DNA 或 mRNA 中与芯片相应基因的突变或表达等情况。它的特点是：一次性检测样本量大、成本相对低、计算机自动分析结果以及快速、准确等，因此适于靶 DNA 或 mRNA 的高通量筛选。

基因芯片应用于靶标筛选的方法主要有两种。一种是获得未知功能的基因并通过芯片技术进行分析。基因芯片可以同时筛选大量的基因并通过数据的对比分析指出可能的新的疾病靶标[10,11]。另一种方法是大规模比较正常组织与疾病组织的表达谱之间的区别。基因芯片技术可以用于鉴别多种疾病中上调（upregulated）和下调（downregulated）的基因。Heller 等[12]利用这种技术比较了风湿性关节炎和肠炎组织与正常组织间的差异，证实这两种炎症疾病的发生与多种基因有关。Wilson 等[13]在基因芯片上点布了与 97% 的结核分枝杆菌基因可读框相对应的 cDNA，检测经过异烟肼作用后对其有耐药性的结核杆菌，发现 5 种编码Ⅱ型脂肪酸合成酶和 fbpC 酶的基因、4 种与生物合成途径无直接相关的基因（efpA、fadE23、fadE24、ahpC）表达增高，这就为将来人工合成新的抗结核药物提供了有用的线索。

在肿瘤研究和诊治方面，Wang 等[14]将 5766 种 cDNA 点布在一张基因芯片上，用它检测正常卵巢组织和卵巢肿瘤之间各基因表达的差异，鉴定了在卵巢癌

变过程中起着重要作用的几种基因，这也为肿瘤的诊断和治疗带来了新的方法和途径。另外还有多种癌症组织[15~17]被人们用这种方法所研究，这些肿瘤基因表达谱的研究对于肿瘤的分类和提供癌症的新型基因靶标提供了可能。

近年来，随着分子生物学的不断发展，越来越多的单基因遗传病的基因被克隆出来。利用基因芯片检测容量大的特点，可根据某个遗传病的特定基因，设计一组包含能检测该基因各外显子上所有可能出现突变的 DNA 序列位点的寡核苷酸探针，制成一张芯片，利用杂交时会出现碱基错配杂交信号减弱的特性，经计算机分析扫描信号，将该基因上的杂合性突变检测出来，并分析出该突变位点在突变后有无 mRNA 表达、表达后翻译的蛋白质功能变化的可能情况，使得一次检测就可以完成对一系列疾病的基因诊断。同时，单张芯片上容纳的寡核苷酸探针越多，一次检测能够诊断的基因遗传病种类也越多。随着基因芯片产量的增加，价格也会逐步下降，使得对基因遗传病的诊断和产前诊断在实际中的应用成为现实。

二、蛋白质芯片

蛋白质芯片又称蛋白质微阵列（protein microarray），它是继基因芯片之后，作为基因芯片功能的补充发展起来的。DNA 基因芯片为描绘不同组织的 mRNA 表达谱提供了可能，并且得到了广泛的应用。蛋白质是生命活动的执行者和体现者，是生物体最主要的结构成分、催化剂和信号转导分子，一个基因在转录时也许以多种 mRNA 形式剪接，而同一种蛋白质可能会以多种形式进行翻译后修饰，蛋白质组中蛋白质的数量可能超过基因组的数量。所以更能反映生物体机能的是蛋白质表达谱，而不是 mRNA 表达谱。蛋白质芯片的发展使得在同一时相分析整个蛋白质组成为可能。

保留色谱和预活化芯片表面是蛋白质芯片的两种常用的分离手段[18]。用于蛋白质芯片表面的色谱主要有：反相色谱、离子色谱（阴离子和阳离子）交换色谱、金属亲和色谱；预活化表面主要是基于蛋白质之间的专一性作用，如：抗原－抗体结合和受体－配基等（见图 15-4）。通过分离，可以将从细胞或体液中分离出的具有某些共同特性的蛋白质富集至蛋白质芯片上用于进一步的分析。

与基因芯片相似，蛋白质芯片就是在载体上点布高密度不同种类的蛋白质，然后再用标记了荧光染料的已知抗体或配体等与待测样本中的抗体或配体一起同芯片上的蛋白质竞争结合，在扫描仪上读出荧光强弱，利用计算机分析计算出待测样本结果。在此基础上进一步发展到对各种蛋白质、抗体以及配体的检测，弥补了基因芯片检测的空缺。一般来说，制作蛋白质芯片及其检测的流程是：首先用噬菌体展示技术制备抗体文库，再将其在芯片载体上布阵；分别制备正常组织和疾病组织的蛋白质组，对蛋白质组进行适当的标记；蛋白质组与芯片上的抗体

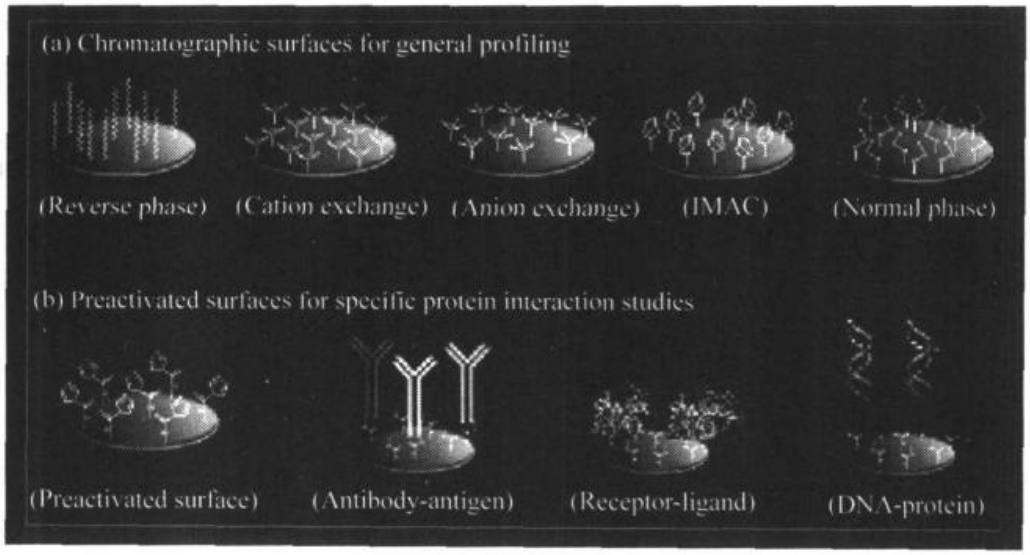


图 15-4　各种形式的蛋白质芯片表面

发生反应，只有与芯片上的抗体发生反应的蛋白质在洗脱后能结合在芯片表面；进行 ELISA 试验，分析试验结果，根据各点的强度进行比较分析。机器人在这一领域的应用使得蛋白芯片在受体-配基相互作用和基因表达的高通量筛选中的应用成为可能。Lueking 等[19]首先利用机械手将基因工程生产的 92 种不同种类的人体蛋白质点布在聚偏二氟乙烯（PVDF）滤片上，用相应的单克隆抗体可以测得很好的结果，其中 ELISA 反应得到的图像经过 CCD（charge-coupled device）照相技术及激光扫描系统等对信号进行检测，整个检测过程完全自动化，真正做到了高效率、低成本，特别适合于蛋白表达的大规模、多种类筛查，提供高通量的配体-受体反应研究手段。Mendoza 等[20]则采用了毛细管系统将单克隆抗体和抗原转移到玻璃表面，每个阵列最多可以有 144 种单独的抗体和抗原。但在高通量模式下，一次可以印刷 96 个阵列，这样就可以同时分析 96 个不同的样品，达到高通量的要求。

对吸附到蛋白质芯片表面的靶蛋白的检测主要有两种方式。除上面提到的蛋白质标记法，样品中的蛋白质预先用荧光物质或同位素等标记，结合到芯片上的蛋白质就会发出特定的信号，用 CCD 及激光扫描系统等对信号进行检测的方法外，还有一种是以质谱技术为基础的直接检测。表面增强激光解析离子化-飞行时间质谱技术（surface enhanced laser desorption/ionization time-of-flight mass spectrometry，SELDI-TOF-MS）在蛋白芯片分析中的应用使得检测的灵敏度大大增强。在点阵的表面衍生传统的色谱分离介质，如反相介质、离子交换介质、固相金属亲和介质和混合模式的介质等，生物分子通过不同的作用与表面上的各种介质结合，因为其作用具有高度专一性，可以得到比标准的层析介质更显著的富集效果，从而在 SELDI-TOF-MS 中获得比普通 MALDI-TOF-MS 信号更强（灵敏度更高）、更均一、重复性更好的谱图，进行蛋白质的相对分子质量和相对定

量分析[21]。与DNA芯片一样，蛋白质芯片同样蕴涵着丰富的信息量，必须利用专门的计算机软件包进行图像分析、结果定量和解释，才能应用于疾病靶标蛋白质的发现与筛选研究。

目前我国在生物芯片研发及生产技术、应用等方面已经取得了很大的成功[22]。世界首例用于肿瘤早期检测的多肿瘤标志物蛋白芯片检测系统，日前已在上海数康生物科技有限公司开发研制问世，经多家医院临床试验和各方专家论证鉴定，国家药品监督管理局已正式为其颁发了生物制品一类新药证书及试生产批文。这一系统是基于生物芯片技术和免疫学技术，通过检测12种肿瘤标志物，可以达到对10种常见肿瘤（原发性肝癌、肺癌、前列腺癌、胰腺癌、胃癌、食道癌、卵巢癌、子宫内膜癌、结/直肠癌、乳腺癌）进行同时检测的目的，可广泛地应用于临床辅助诊断、体检普查等领域。

第二节　药物的高通量筛选体系

综合运用蛋白质组技术等筛选得到疾病靶标蛋白质以后，可以合理设计针对疾病靶标特点的先导化合物及其活性类似物。要想得到临床药物，还需要对其进行大规模活性筛选。在人类基因组计划进行的近几年中，药物筛选有了很大发展，现代药物筛选包括以下几个层次：① 计算机模拟药物筛选和计算机辅助设计。运用计算机技术，以药物靶标分子三维结构和蛋白质晶体结构为基础，对含有大量化合物结构的数据库进行模拟“筛选”，迅速高效地发现先导化合物及其新用途。② 分子水平的药物筛选模型。利用生命活动中发挥重要作用的基因位点、受体、酶、离子通道、核酸等生物大分子作为药物筛选的作用靶点进行分子水平筛选，其中，基因芯片和蛋白质芯片将被广泛应用。③ 细胞水平的药物筛选模型。利用离体培养细胞株、转基因细胞株进行药物作用的筛选或鉴定。④ 动物个体水平的药物筛选模型。动物试验作为药物在人体临床前的鉴定模型，转基因和基因剔除动物模型的平台建立，为药物筛选和基因治疗提供了实验模型。⑤ 人体临床筛选模型。作为药物上市前的最后筛选。

采用分子水平的筛选模型已被普遍用于除计算机模拟药物筛选外的第一步初筛，其优点是特异性强、灵敏度高、微量快速。蛋白质组学技术应用于筛选模型构建中的一个最大的优势就是更清楚、更详尽地阐明分子药理机制，提供更为有效、合理的药理模型。现在蛋白质组学的研究结果表明，蛋白质基因表达的调控成了真正的药物作用机制。在药物发现阶段，大量合成的有机化合物和分离得到的天然产物有效成分，在有效的药理模型上利用自动化的筛选技术对其进行随机筛选，发现具有进一步开发价值的化合物，称为先导化合物（lead compound）。药物发现阶段对创新药物的研究具有决定性的意义，筛选效率的提高将大大缩短新药发现的周期。蛋白质组学技术除能够促进靶点的发现外，还有助于筛选模型

的建立。

近年来，随着科学技术的进步，如分子生物学、分子病理学、分子药理学、自动化技术（特别是机器人的应用）和微电子技术等学科的发展，在新药研究中形成了药物发现过程的新方法，出现了高通量筛选（high throughput screening，HTS）体系，它有机地将组合化学、基因组研究、生物信息和自动化仪器、机器人等先进技术进行组合，形成一种发现新药的新程序[23]。高通量筛选是药物产业化过程中的一个关键环节，它是测试标靶相对于批量化合物的结合能力或生物学活性。被测试化合物可能作为标靶的抑制物、天然受体结合位点的竞争物、细胞内受体介导过程的激活物或拮抗物等。制药公司一般一次要筛选 100 000～300 000个以上的化合物，从中确定 100～300 个候选物，最后，一至两个化合物能成为先导化合物。

在 HTS 发展之前的 90 年代初期，采用传统的方法，一个实验室借助 20 余种药物作用靶位，一年内可以筛选 75 000 个样品；到了 1997 年 HTS 发展的初期，采用 100 余种靶位，每年可筛选 1 000 000 个样品；而到 1999 年，由于 HTS 的进一步完善，每天的筛选量就高达 100 000 种化合物[24]，因此称为超高通量筛选（ultrahigh-throughput screening）。这种新的飞跃，将大大加速新药发现的速度。

HTS 系统应用和推广的关键在于必须具备相应的新药筛选模型。近年来以基因表达产物为模型建立的新药筛选模型为 HTS 的应用提供了有利条件。如根据人类重要疾病的种类，以每种疾病可能涉及的基因及每种基因产物在致病过程中可能与之相关的蛋白相互作用数目计算，预计基因组研究的结果可提供5000～10 000 种分子药靶，靶位数目将超过以往的 20 余倍。

一、高通量筛选

高通量筛选系统主要由五个部分组成：① 自动化操作系统；② 高灵敏度检测系统；③ 分子、细胞水平的高特异性体外筛选模型；④ 被筛样品库；⑤ 数据采集传输处理系统。除此之外，尚有一些辅助系统配合高通量药物筛选，如计算机辅助设计、组合化学、高效天然化合物提取方法等。用这些现代方法进行药物筛选，真正实现了快速、大规模筛选和一药多筛。

自动化操作系统是高通量筛选的基础。因为每天要筛选检测成千上万化合物的样品，工作重复而枯燥，人工操作显然满足不了准确、快速的要求。自动化操作系统本身主要由计算机及其操作软件、自动化控制设备、温孵、离心（加样和冲洗）等设备和堆栈（hotel）[25]等部分组成。它可以代替人工完成自动加样、转移、洗脱、混合、温孵和检测等操作，减少人工操作所造成的误差。

高灵敏度、快速的检测系统是高通量筛选的关键。这种筛选通常在 96 孔

(384 孔甚至更多孔) 的微孔板上进行，使用的样品量小，要求仪器灵敏度高、满足快速检测的需要。用于 HTS 中的检测方法，大多以 ELISA 方法为基础，其目的在于确定配基和受体的相互作用。目前多采用荧光或酶标比色方法代替同位素标记法。

分子和细胞水平的高特异性体外筛选模型为高通量的自动化筛选奠定了基础。与整体动物和组织器官水平的筛选模型相比，它具有材料用量少、药物作用机制明确和可实现大规模筛选等特点。与传统模型相比，以基因表达产物为靶位建立的药物筛选模型具有以下突出优点：① 靶位专一、特异性强，因此假阳性率低；② 通过基因克隆扩增可以获得足够量的某些因含量低而无法提取的靶位物质；③ 与采用动物及其组织提取物的方法比较，费用明显降低；④ 所建立的模型适用于 HTS 全自动快速大量筛选的新体系。如何发现更多新的药靶，并在此基础上建立各种特异性强的新的筛选模型，是促进新药筛选研究的关键因素。随着人类和模式生物基因组及后基因组研究的飞速发展，其结果必将极大地推动新药的相关研究。

采用 HTS 进行筛选的样品主要有组合化学库、天然化合物、化学合成化合物、反义核酸、肽核酸、可溶性蛋白等，生物芯片技术也可用于 HTS 药物筛选研究。

数据采集传输处理系统实际上是对大规模、高通量筛选结果的总结。一方面，它负责将筛选的结果采集和传输到数据库中；另一方面，它还将对数据进行处理和管理。数据库系统需要对筛选化合物样品的理化性质、生物活性信息进行存储管理。同时，通过对同一模型的呈现阳性反应的化合物结构进行分析，从而为进一步的药物设计提供参考，达到大规模高通量的筛选目的。

如何发现更多新的药靶，并在此基础上建立各种特异性强的新的筛选模型，是促进新药筛选研究的关键因素。人类基因组计划的完成及后基因组研究的飞速发展，必将极大地推动新药的相关研究。

二、计算机虚拟筛选[26,27]

计算机虚拟筛选 (in *silico* screening/ virtual screening) 主要用于限定 (减少) 提交生物学筛选的化合物数量和结构特征。它是综合计算机辅助药物分子设计等多学科的一种筛选优筛 (focussed screening) 策略。它覆盖了计算机辅助药物分子设计的大部分研究领域，从组合化学库的合理设计到三维结构化学关系数据库的自动产生；从库中分子结构相似性分类到化合物是否具备药物类似性的快速预测；从对系列化合物三维定量构效关系研究与药效团的构建到药物分子与受体相互作用的精确计算机模拟。其目的是在大量化合物样品库中挑选出能代表分子多样性和药物类似性的分子作为候选先导化合物进行生物学筛选，从而大大减

少药物开发的消耗和缩短药物开发的周期。

第三节　药物作用的监测评价体系

经过大规模筛选后的先导化合物进入药物开发阶段。这一阶段包括药效学研究、药代动力学研究、安全性评价和临床试验等步骤。这一过程最大的特点是耗时又耗资，且常常是幸存者寥寥无几。尤其是在临床试验中被淘汰出局的为多数。分析其原因是缺乏对药物的毒理机制深入了解和缺乏有效的临床前安全性评价指标。蛋白质组学的出现及时有效地弥补上了这块空白。

一、临床前安全性评价

人们越来越认识到蛋白质组学在药物毒理机制研究中的重要性。蛋白质组数据库能根据蛋白质点的强度反映一种疾病的过程变化或对一组药物反应的变化。这样，蛋白质可以用作药物反应或疾病的替代标志。环孢素 A 造成严重的肾毒性原因的探明是蛋白质组学在药物毒理机制研究中成功应用的一个很好的例子[28]。环孢素 A 是一种应用广泛的免疫抑制剂，被成功地用于器官移植手术中和自身免疫疾病的治疗中，但应用中也伴随着许多副作用，其中最值得注意的是肾毒性。用环孢素 A 处治后最明显的结果是，部分蛋白点消失，这些蛋白点后来从大鼠肾脏蛋白质模式中被鉴定为 calbindin-D 的 28 kDa 蛋白。使用酶联免疫吸收剂分析的进一步研究证实，大鼠肾脏中 calbindin-D 的 28 kDa 蛋白量持续减少。这种减少伴随着尿钙排泄物和皮层脊髓管内石灰化的增加。就像以前提出的那样，如果 calbindin-D 28 kDa 蛋白扮演钙运输的促进器和钙缓冲器的角色，观察到的效应就能得到解释。环孢素 A 毒性标志物的鉴定，为其毒效提供了可靠的说明，这是蛋白质组与其他的技术结合的成功应用结果。第二个生物标记蛋白质成功鉴别的例子来自膀胱鳞片状细胞癌进程的长期研究[29]。在此研究中，可以监测定义蛋白质标志物和癌损害的分化程度。

因此，无怪乎 Aicher 等[30]这样评论蛋白质组：“鉴于目前蛋白质组相关技术已取得的长足进步，人们有信心期待大规模蛋白质组学将成为进行临床安全性评价的重要工具。”

二、临床试验

利用蛋白质组学可以检测到疾病相关蛋白质，这些蛋白质实际上就是指示疾病发生的分子标志物。因而，这些分子信号将来有可能被用作在临床试验中监测受试者用药反应的参考。同时也可以这些分子信号为依据进行最适临床受试者的

挑选。

三、生物技术药物的开发

生物技术药物包括基因工程重组蛋白质药物、寡核苷酸药物、基因工程抗体药物、DNA 药物等。生物技术药物发展的主要动力来源是基因组和蛋白质组的研究。其中最令人注目的进展是抗体药物。

抗体是机体抵抗病原体的主要防御系统。它们是由免疫系统自主产生的蛋白质。每个抗体都可特异性地识别作为抗原靶分子的特征，哪怕是极其微小的差别。目前，大多数抗体药物都是鼠人嵌合体抗体或人源化抗体。由于传统的产生抗体方法是基于鼠的免疫机制，来源是鼠类，因而临床易发生免疫排斥反应。针对这一问题，CAT 生物技术公司正在利用蛋白质组学独特的技术体系，以及组合噬菌体抗体文库（combinatorial phage antibody librarys），加上后基因组学的一些技术，开发能够成为今后人类治疗药物的人抗体。

随着后基因组时代的到来和科学技术尤其是蛋白组学、生物信息学等的发展，蛋白组学在药物开发中的作用会越来越明显，通过蛋白质组学开发出的药物也会逐渐深入到人们的生活中去。

参 考 文 献

1 肖著. 靶向给药与新药开发. 见：21 世纪分子医药学前沿. 天津：天津科学出版社. 2001，210－213

2 Nuttall M E. Drug discovery and target validation. Cell Tissues Organs, 2001, 169: 265－271

3 Dean P M, Zander E D, Bailey D S. Industrial-scale, genomics-based drug design and discovery. Trends in Biotech, 2001, 19: 288－292

4 Martinez F D, Grves P E, Baldini M, et al. Association between genetic polymorphisms of the beta 2-adrenoceptor and albuterolin children with and without a history of wheeling. J Clin Invest, 1999, 100: 3184

5 Müllner S, Neumann T, Lottspeich F. Proteomics — a new way for drug target dicovery. Arzneim. -Forsch. /Drug Res, 1998, 48: 93－95

6 Schena M, Heller R A, Theriault T P, Konrad K, Lachenmeier E, Davis R W. Microarrays: biotechnology's discovery platform for functional genomics. Trends Biotechnol, 1998, 16 (7): 301－306

7 Blohm D H, Guiseppi-Elie A. New developments in microarry technology. Curr Opin Biotechnol, 2001, 12: 41－47

8 Okamoto T, Suzuki T, Yamamoto N. Microarray fabrication with covalent attachment of DNA using bubble jet technology. Nat Biotechnol, 2000, 18: 438－441

9 Singh-Gasson S, Green R D, Yue Y, Nelson C, Blattner F, Sussman M R, Cerrina F. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. Nat Biotechnol, 1999, 17: 974－978

10 Chee M, Yang R, Hubbell E, Berno A, Huang X C, Stern D, Winkler J, Lockhart D J, Morris M S, Fodor S P A. Accessing genetic information with high-density DNA arrays. Science, 1996, 274: 610－614

11 Lashkari D A, DeRisi J L, McCusker J H, Namath A F, Gentile C, Hwang S Y, Brown P O, Davis R W.

Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc Natl Acad Sci. USA, 1997, 94: 13057－13062

12 Heller R A, Schena M, Chai A, Shalon D, Bedilion T, Gilmore J, Woolley D E, Davis R W. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. Proc Natl Acad Sci USA, 94: 2150－2155

13 Wilson M, DeRisi J, Kristensen H H, Imboden P, Rane S, Brown P O, Schoolnik G K. Exploring drug-induced alterations in gene expression in Mycobacterium tuberculosis by microarray hybridization. Proc Natl Acad Sci USA, 1999, 96: 12833－12838

14 Wang K, Gan L, Jeffery E, Gayle M, Gown A M, Skelly M, Nelson P S, Ng W V, Schummer M, Hood L, Mulligan J. Monitoring gene expression profile changes in ovarian carcinomas using cDNA microarray. Gene, 1999, 229: 101－108

15 Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A, Sampas N, Dougherty E, Wang E, Marincola F, Gooden C, Lueders J, Glatfelter A, Pollock P, Carpten J, Gillanders E, Leja D, Dietrich K, Beaudry C, Berens M, Alberts D, Sondak V. Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature, 2000, 406: 536～540

16 Xu J, Stolk J A, Zhang X, Silva S J, Houghton R L, Matsumura M, Vedvick T S, Leslie K B, Badaro R, Reed S G. Identification of differentially expressed genes in human prostate cancer using subtraction and microarray. Cancer Res, 2000, 60: 1677－1682

17 Perou C M, Jeffrey S S, van de Rijn M, Rees C A, Eisen M B, Ross D T, Pergamenschikov A, Williams C F, Zhu S X, Lee J C, Lashkari D, Shalon D, Brown P O, Botstein D. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci USA, 1999, 96: 9212－9217

18 Weinberger S R, Dalmasso E A, Fung E T. Current achievements using ProteinChip® Array technology. Current Opinion in Chemical Biology, 2001, 6: 86－91

19 Lueking A, Horn M, Eickhoff H, Bussow K, Lehrach H, Walter G. Protein microarrays for gene expression and antibody screening. Anal Biochem, 1999, 270: 103－111

20 Mendoza L G, McQuary P, Mongan A, Gangadharan R, Brignac S, Eggers M. High-throughput microarray-based enzyme- linked immunosorbent assay (ELISA). Biotechniques, 1999, 27: 778－788

21 Merchant M, Weinberger S R. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. Electrophoresis, 2000, 21: 1164～1177

22 http: //www. health-digit. com/gb/products. htm

23 John G H, Martyn B. The chemical-biological interface: developments in automated and miniaturised screening technology. Curr Opin Biotechnol, 1997, 8: 734－740

24 Alan D. Drug screening-beyond the bottleneck. Nature Biotechnol, 1999, 17: 859－863

25 黄家学，胡娟娟，杜冠华. 化合物药物活性的高通量筛选. 中国药理学通报，1999，15 (5)：401～403

26 Gane P, Dean P M. Recent advances in structure-based rational drug design. Curr Opin Struct Biol, 2000, 10: 401－444

27 Valler M J, Green D. Diversity screening versus focused screening in drug discovery. Drug Discov Today, 2000, 5: 286－293

28 Evers S, Gray C P. Application of proteome analysis to drug development and toxicology. In: Pennington S, Dunn M J, ed. Proteomics: From Protein Sequence to Function. New York: Springer-Verlag New York Inc. 2001. 225－236

29 Celis J E, Ostergaard M, Jensen N A, Gromova I, Rasmussen H H, Gromov P. Human and mouse proteom-

ic databases: novel resources in the protein universe. FEBS Lett, 1998, 430: 64～72

30 Aicher L, Wahl D, Arce A, Grenet O, Steiner S. New insights into cyclosporine A nephrotoxicity by proteome analysis. Electrophoresis, 1998, 19: 1998～2003

（刘尚义）